Furthering our Understanding of Heterogeneous Ice Nucleation with Molecular Simulation

Stephen James Cox

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Chemistry

UCL

December 11, 2014

I, Stephen James Cox, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

Ice formation is arguably the most common phase transition on the planet and almost always occurs heterogeneously. Despite the importance of ice formation to the climate, medical and geological sciences, as well as the food and transport industries, a clear understanding of how the properties of a material affect its ability to nucleate ice has remained elusive. This has prevented the rational design of new materials to either inhibit or promote ice nucleation. In this thesis, a wide variety of computational techniques are used to try and further our understanding of heterogeneous ice nucleation. This includes: testing long established theories; investigating ice formation in the presence of a known ice nucleating agent; using simplified model surfaces to elucidate the underlying molecular mechanisms of heterogeneous ice nucleation (and design new ice nucleating agents *in silico*); and developing transition path sampling techniques to look at some of the fundamental aspects of homogeneous nucleation. The accuracy of commonly used approximations to define the potential energy surface of a closely related system, methane hydrate, is also investigated.

Publications

The work presented in this thesis has led to the following publications:

1. Stephen J. Cox, Shawn M. Kathmann, John A. Purton, Michael J. Gillan and Angelos Michaelides, *Non-hexagonal ice at hexagonal surfaces: the role of lattice mismatch*, Phys. Chem. Chem. Phys. **14**, 7944 (2012)

2. Stephen J. Cox, Zamaan Raza, Shawn M. Kathmann, Ben Slater and Angelos Michaelides, *The microscopic features of heterogeneous ice nucleation may affect the macroscopic morphology of atmospheric ice crystals*, Faraday Discuss. **167**, 389 (2013)

    - See also: Chandler *et al.*, *General discussion*, Faraday Discuss. **167**, 455 (2013)

3. Stephen J. Cox, Michael D. Towler, Dario Alfè and Angelos Michaelides, *Benchmarking the performance of density functional theory and point charge force fields in their description of sI methane hydrate against diffusion Monte Carlo*, J. Chem. Phys. **140**, 174703 (2014)

4. Stephen J. Cox, Shawn M. Kathmann, Ben Slater and Angelos Michaelides, *Controlling ice nucleation through surface hydrophilicity* (submitted)

Acknowledgements

I have thoroughly enjoyed my time in London. I am indebted to my supervisor, Angelos Michaelides, for his guidance throughout my PhD. I am also grateful to Shawn Kathmann and his colleagues at Pacific Northwest National Laboratory for their input into this project. All group members, past and present, are thanked for making the office such a fun environment to work in. In particular, I would like to mention my direct peers Gabriele Tocci and Gabriella Graziano, who have been ever-present throughout my time in the group. Brent Walker is also acknowledged for his willingness to help with aspects of coding even after he moved to the private sector. I am fortunate to have collaborated with many people over the course of my PhD. Ben Slater's contributions to the work presented in Chapters 4 and 5 are duly acknowledged. I am also grateful for collaborations at various points of the project with Mike Gillan, Dario Alfè, John Purton, Zamaan Raza, David Limmer, David Chandler and Neal Skipper. David Limmer is also thanked for sharing a version of his TPS code, upon which the code used in Chapter 6 was based. I would like to thank my parents for their support, especially over the course of my undergraduate studies, which enabled me to do this PhD. Finally, I thank my wife, Julie, whose love and support have been of inestimable importance.

Contents

A CD-ROM with supplementary material and an electronic copy of this thesis is enclosed on the inside of the back cover.

List of Figures

List of Tables

Chapter 1

Introduction

The formation of ice is one of the most common phase transitions on the planet and is important to a wide variety of fields such as the climate, biological and geological sciences, as well as the food, energy and transport industries. The airline industry, for example, is in a constant battle to prevent ice formation on the wings and in the fuel lines of aircraft. Failure to do so can result in disaster, such as in 2008 when a British Airways flight from Beijing fell some 300 m short of the runway at London Heathrow due to ice formation on the plane's fuel/oil heat exchanger. In the atmosphere, most cloud formation and precipitation is dependent upon the formation of ice, which generally occurs at the surfaces of aerosol particles, and the amount of ice in clouds is a determining factor of their properties, including the overall levels of solar radiation and heat reflected and adsorbed by the atmosphere [1]. Due to major gaps in our understanding of the fundamental processes that govern ice formation in clouds, it is still unknown if it has a net cooling or heating effect on the climate, as highlighted in the 2013 Intergovernmental Panel on Climate Change Report [2]. Ice formation is also becoming ever more relevant to the energy industry. As readily available oil reserves are depleted, energy companies are required to drill in harsher conditions (lower temperature and higher pressure), making ice and gas hydrate[1] formation more probable. Not only is effective inhibition of ice and hydrate formation desirable

[1]Gas hydrates are crystalline ice-like materials containing dissolved gas molecules such as CH_4 or CO_2. They form under conditions of low temperature and high pressure.

from an economic point of view, but failure of current inhibitor technologies can have catastrophic consequences [3, 4]. The urgency with which we need to understand climate change and our increasing dependence upon inexpensive energy resources and the facile transportation of people and goods, all make the molecular level understanding of ice formation not only interesting, but essential.

Given the ubiquity of ice formation, one might reasonably expect that the underlying physical principles would already be understood. Certainly ice formation has been widely studied, but to claim that we have even a half-decent understanding is a falsehood (see e.g. References [5] and [6]). One undeniable truth is that almost all ice formation occurs *heterogeneously*, that is, in the presence of impurity particles such as mineral dust [7], soot [8–10], pollen [11–13] or bacteria [14]. In fact, it is remarkably easy for water to remain in a metastable 'supercooled' liquid state to temperatures as low as *ca.* −37℃ [7]. It is known empirically that different materials nucleate ice with different efficiencies, as highlighted by Figure 1.1, which shows the variation in the density of potential atmospheric ice nuclei with temperature.[2] Here we can see that, not only are the atmospheric concentrations of different ice nucleating agents (INAs) markedly varied, but so too are the temperatures at which they become active; bacterial INAs, for example, become significantly active between −2 and −12℃, whereas soot only becomes active below −18℃ (although in much higher concentrations).

In general terms, we can explain why materials enhance ice nucleation through the concept of *free energy*. What is meant by 'free energy' will be formalised in the next chapter, but for the moment, it is sufficient to state that the free energy of a system determines its thermodynamic stability, with states of lower free energy preferable over those of higher free energy. At ambient pressure, the free energy of liquid water and crystalline ice are the same at the equilibrium melting temperature of 0℃ (they are said to 'coexist'). Upon cooling the temperature further, the crystal becomes lower in free energy than the liquid. The liquid is not unstable, however, but metastable i.e. the collection of microstates that define the liquid form a local minimum on the free energy surface (referred

[2]The number of potential ice nuclei is calculated by assuming that each nucleating particle (with a known temperature dependent active site density for ice nucleation) is singly immersed within one water droplet [7].

Figure 1.1: Variation in the density of potential atmospheric ice nuclei with temperature. Not only are the concentrations of different materials markedly varied, but so too are the temperatures at which they become active ice nucleating agents. Taken from Reference [7].

to as the 'liquid basin'). As such, there is a barrier on the free energy surface separating the liquid and crystal basins that the system must overcome if it is to crystallise. A schematic of such a free energy surface is presented in Figure 1.2. It is the presence of this free energy barrier that gives rise to the phenomenon of supercooled water. The effect of a good INA is to provide an alternative path (or set of paths) connecting the liquid and crystal basins that has a lower free energy barrier.

So far we have established that most ice nucleation occurs heterogeneously and that different INAs facilitate ice formation with different efficacies. To explain this, the concept of 'free energy' was invoked and and it was stated that good INAs lower the free energy barrier separating the liquid and crystal basins. Within this picture, then, the more potent the INA, the greater the reduction in the free energy barrier. What we do not know is how INAs reduce the free energy barrier, and which are the relevant properties of an INA that determine its ice nucleating ability. It is the last point that forms the basis of this thesis.

Although a significant amount of high quality experimental work has been performed to investigate ice nucleation, the simultaneous temporal and spatial resolution that reveal the underlying molecular mechanisms has remained elusive

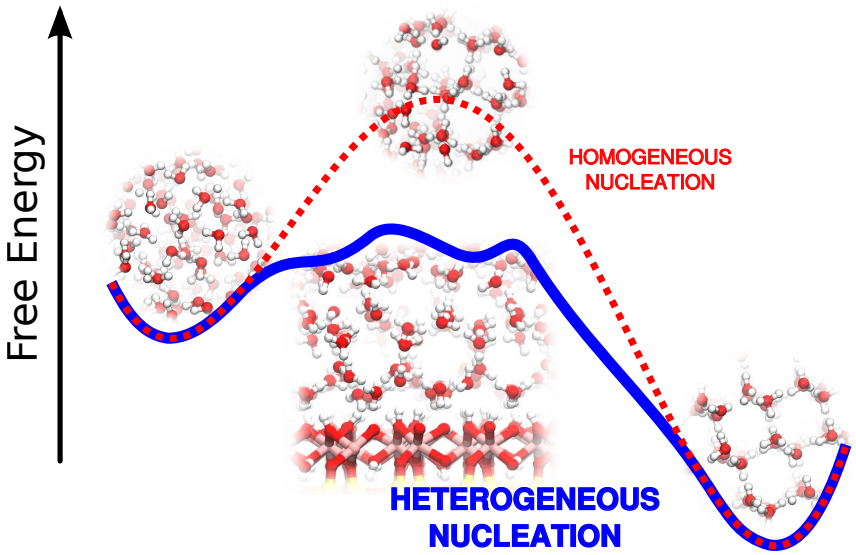

Figure 1.2: Schematic free energy profiles for homogeneous and heterogeneous ice nucleation. The liquid and crystal free energy basins are depicted on the left and right, respectively. The free energy of the crystal basin is lower than that of the liquid. As the liquid transforms to ice, it must overcome a free energy barrier: a good INA will reduce this free energy barrier compared to homogeneous nucleation, making it easier for the liquid to freeze.

[15].[3] Atmospheric chemistry and surface science have probably contributed the most relevant experimental data to this topic. The two fields, however, probe vastly different length-scales and conditions. Generally speaking, atmospheric chemists are concerned with trying to understand the effects of 'real' particles (such as dust from the Sahara desert) on ice nucleation, whereas surface scientists operate under well-controlled conditions with well characterised surfaces. The advantage of the atmospheric chemists' approach is that they actually investigate the problem at hand – how different particles affect ice nucleation. The disadvantage is that little in the way of why the particles behave how they do is revealed; the purity of the samples is often the highest level of molecular detail sought. This approach often produces useful results: recent work in which pure samples of mineral dust were examined elucidated that K-feldspar is a much more potent INA than clay minerals, for example [17]. Surface science experiments, on the other hand, give direct information on how the molecular properties of a surface

[3] Actually, it should be noted that very recent advances in X-ray scattering techniques have allowed real-time monitoring of homogeneous nucleation in micron sized water droplets [16]. Such techniques could be used to investigate the molecular mechanisms of heterogeneous ice nucleation.

affect the behaviour of interfacial water molecules. Such experiments have shown that the behaviour of water at solid surfaces is complex, with water forming a wide variety of structures, even at seemingly similar surfaces (for a recent review, see Reference [18]). The problem associated with this approach is that the direct consequences for ice nucleation are lost. Computer simulation techniques therefore provide an appealing way to study heterogeneous ice nucleation, as they enable us to 'see' how water molecules behave at surfaces, often under conditions inaccessible to surface science experiments. Of course difficulties are encountered and approximations have to be made when modelling molecular processes, and we will discuss these throughout, but by the end of the thesis, it is hoped that the reader is convinced that computer simulation is a powerful tool in the study of nucleation processes.

The thesis is organised as follows: in the next chapter, some of the background theory will be discussed; in Chapter 3, work is presented in which grand canonical Monte Carlo and idealised surfaces were used to study the role of surface lattice constant [19]; Chapter 4 sees results from the first all-atom simulations of heterogeneous ice nucleation [20]; in Chapter 5, a coarse-grained potential for water is used to investigate ice nucleation in the presence of nanoparticles of varying shape and hydrophilicity, whilst in Chapter 6, transition path sampling techniques are used to look at the role of particle mobility in homogeneous nucleation; Chapter 7 is the final results chapter in which we discuss the accuracy of commonly used density functional theory (DFT) approaches in describing bulk sI methane hydrate [21]; we finish with a summary and outlook in Chapter 8.

To conclude these introductory remarks, we will discuss an example of how powerful a fundamental understanding of molecular processes can be. In the field of chemical catalysis, DFT calculations are fast and often accurate enough to allow the complete kinetics of complex catalytic reactions (i.e. reaction barriers, energies and entropies) to be evaluated [22]. Nevertheless, it is impractical to map the entire chemical kinetics for a huge number of different catalysts for many different reactions. Instead, the insight gained from the DFT calculations can be used to determine which properties at the molecular level determine the macroscopic kinetics. For example, DFT calculations of the methanation reaction (CO +

$3\,H_2 \rightarrow CH_4 + H_2O$) identified the active sites on the catalyst as steps or edges [23]. Comparison of different metal surfaces as the catalyst then showed that the barriers for CO activation, along with CH_4 and H_2O formation were found to scale essentially linearly with the dissociation energy of CO on the surface: as the CO dissociation energy increases, so too does the CO activation barrier, but the CH_4 and H_2O formation barriers decrease. The Sabatier principle therefore holds,[4] and an optimal value for the CO dissociation energy exists. As the CO dissociation energy is a computationally tractable quantity, a computational screening of a series of binary alloys found that NiFe alloys have a high catalytic activity at low cost [23]. It is hoped that the work in this thesis can act as a platform from which future studies can identify computationally tractable 'descriptors' (like the CO dissociation energy) that can be used to predict the ice nucleating ability of existing materials, as well as help design new materials to control ice formation.

[4]The Sabatier principle states that a good catalyst should bind an adsorbate neither too strongly nor too weakly. If the catalyst-adsorbate interaction is too weak, then the adsorbate will fail to bind to the catalyst. If the interaction is too strong, the catalyst will either get blocked by the adsorbate or fail to release the products.

Chapter 2

Background Theory

Due to its complexity, a variety of techniques have been used in attempting to elucidate the underlying molecular mechanisms of heterogeneous ice nucleation. These include grand canonical Monte Carlo, molecular dynamics (with both atomistic and coarse grained potentials), transition path sampling and DFT. Rather than discussing the details of each of these methods here, where they would be far removed from their corresponding results, each chapter will instead begin with a section outlining the method appropriate to it. In the current chapter, we will limit ourselves to discussing some of the general concepts that will be useful throughout the rest of the thesis.

2.1 Statistical Mechanics

2.1.1 Introduction

Statistical Mechanics is a theoretical approach that allows us to relate the microscopic behaviour of many body systems to their macroscopic observable behaviour. Statistical Mechanics can provide analytically exact equations to compute the properties of simple systems (e.g. the ideal gas or Einstein crystal). For more complex systems, such as water, it provides the framework for the computational techniques that will be used throughout the thesis. Many excellent texts exist that give a thorough introduction to Statistical Mechanics (see e.g. [24–27]) and only a brief overview will be given here. Throughout the thesis, the assumption

will be made that classical mechanics can be used to describe the motion of atoms and molecules in a many-body system. In the spirit of Reference [25], however, we need to borrow one concept from quantum mechanics: that a quantum mechanical system can be found in different states. Considering states that are eigenstates of the system's Hamiltonian $\hat{H}$, for any state $|i\rangle$ we have $\hat{H}\,|i\rangle = E_i\,|i\rangle$, where E_i is the energy of state $|i\rangle$.

2.1.2 The Ergodic Principle, Entropy and Temperature

For macroscopic systems (i.e. systems of $\mathcal{O}(10^{23})$ particles), the degeneracy of the energy levels is a beyond-astronomically large number. For a system of N particles in a volume V, the number of states with energy E is denoted as $\Omega(E, V, N)$. The ergodic principle states that:

> A system of fixed N, V and E is equally likely to be found in any of its $\Omega(E)$ states.

If we now consider two systems that are able to exchange energy such that the total energy of the two systems $E = E_1 + E_2$ is constant, this can be achieved by distributing the energy amongst the two systems in many ways. Furthermore, we stipulate that the systems cannot exchange particles or do work on each other. For a given E_1, the total number of allowed states is $\Omega(E) = \Omega_1(E_1) \times \Omega_2(E_2)$. We note that the total number of states is the product of the degeneracies in each subsystem. By taking the natural logarithm, we can obtain an extensive quantity (i.e. a quantity that scales linearly with system size):

$$\ln \Omega(E_1, E - E_1) = \ln \Omega_1(E_1) + \ln \Omega_2(E - E_1). \tag{2.1}$$

As the two systems do not exchange particles or do work on each other, we deduce that the exchange in energy corresponds to a heat transfer process. We are interested in the most likely way that this energy is distributed between the two systems. Although every energy state of the total system is equally likely, the number of energy levels that correspond to a given distribution between systems 1 and 2 depends highly on E_1. To find the most likely value, we therefore need to

find the value of E_1 that maximises $\ln \Omega(E_1, E - E_1)$:

$$\left(\frac{\partial \ln \Omega(E_1, E - E_1)}{\partial E_1}\right)_{N,V,E} = 0, \tag{2.2}$$

which implies that:

$$\left(\frac{\partial \ln \Omega_1(E_1)}{\partial E_1}\right)_{N_1,V_1} = \left(\frac{\partial \ln \Omega_2(E_2)}{\partial E_2}\right)_{N_2,V_2}. \tag{2.3}$$

If we define:

$$\beta(E, V, N) \equiv \left(\frac{\partial \ln \Omega(E, V)}{\partial E}\right)_{N,V}, \tag{2.4}$$

Equation 2.3 can be written as:

$$\beta(E_1, V_1, N_1) = \beta(E_2, V_2, N_2). \tag{2.5}$$

If we now conduct a thought experiment whereby we initially set $E_1 = E$, there will be a flow of heat from system 1 to system 2 until Equation 2.5 is satisfied. Once this has happened, there will be no net flow of energy between systems 1 and 2, and Equation 2.5 is therefore equivalent to the statement that systems 1 and 2 must have the same temperature at thermal equilibrium. Like the thermodynamics entropy S, $\ln(\Omega)$ is an extensive quantity that is maximum at thermal equilibrium. We therefore consider as a definition of the entropy:

$$S(E, V, N) \equiv k_{\mathrm{B}} \ln \Omega(E, V, N), \tag{2.6}$$

where k_{B} is Boltzmann's constant.[1] From Thermodynamics we know that the temperature T is related to the entropy by $1/T = \left(\frac{\partial S}{\partial E}\right)_{N,V}$. Combining Equations 2.4 and 2.6, we establish:

$$\beta = \frac{1}{k_{\mathrm{B}} T}. \tag{2.7}$$

[1] $k_{\mathrm{B}} = 1.38\ldots \times 10^{-23}\,\mathrm{J/K}$. This fundamental constant cannot be derived, and is determined by comparison to experiment.

2.1.3 The Canonical Ensemble and Free Energy

In the above discussion, we considered a total system in which the number of particles N, the volume V and the energy E were held constant. This is known as the *microcanonical ensemble*. It is also useful to consider systems in which, rather than the energy, the temperature T is fixed. This is known as the *canonical ensemble*. We consider a system of interest that is in thermal equilibrium with a large heat-bath. The total system is closed such that the total energy is $E = E_{\text{sys}} + E_{\text{bath}}$ (the subscripts 'sys' and 'bath' correspond to the system and heat-bath, respectively). We now prepare the system in one definite state $|i\rangle$ with energy $E_{\text{sys}}^{(i)}$. The number of states accessible to the total system is thus $\Omega_{\text{bath}}(E_{\text{bath}}) = \Omega_{\text{bath}}(E - E_{\text{sys}}^{(i)})$. By invoking the ergodic hypothesis, we can therefore write the probability of finding the system in state $|i\rangle$:

$$\begin{aligned} P(E_{\text{sys}}^{(i)}) &= \frac{\Omega_{\text{bath}}(E - E_{\text{sys}}^{(i)})}{\sum_i \Omega_{\text{bath}}(E - E_{\text{sys}}^{(i)})} \\ &= \frac{\exp\left(\ln \Omega_{\text{bath}}(E - E_{\text{sys}}^{(i)})\right)}{\sum_i \exp\left(\ln \Omega_{\text{bath}}(E - E_{\text{sys}}^{(i)})\right)}. \end{aligned} \tag{2.8}$$

As the heat-bath is large, $E_{\text{sys}}^{(i)} \ll E$ and we can expand $\ln \Omega_{\text{bath}}(E - E_{\text{sys}}^{(i)})$ around $E_{\text{sys}}^{(i)} = 0$:

$$\ln \Omega_{\text{bath}}(E - E_{\text{sys}}^{(i)}) = \ln \Omega_{\text{bath}}(E) - E_{\text{sys}}^{(i)} \frac{\partial \ln \Omega_{\text{bath}}(E)}{\partial E} + \dots. \tag{2.9}$$

Thus, from Equation 2.4:

$$\ln \Omega_{\text{bath}}(E - E_{\text{sys}}^{(i)}) \approx \ln \Omega_{\text{bath}}(E) - \beta E_{\text{sys}}^{(i)}, \tag{2.10}$$

and combining Equations 2.8 and 2.10 we obtain the Boltzmann distribution:

$$\begin{aligned} P(E_{\text{sys}}^{(i)}) &= \frac{\exp\left(-\beta E_{\text{sys}}^{(i)}\right)}{\sum_i \exp\left(-\beta E_{\text{sys}}^{(i)}\right)} \\ &= Q(\beta, V, N)^{-1} \exp\left(-\beta E_{\text{sys}}^{(i)}\right), \end{aligned} \tag{2.11}$$

where in the last line we have defined the *canonical partition function* $Q(\beta, V, N) \equiv \sum_i \exp\left(-\beta E_{\text{sys}}^{(i)}\right)$. Equation 2.11 allows us to compute quantities such as the average energy of the system:

$$\begin{aligned}\langle E_{\text{sys}} \rangle &= \sum_i E_{\text{sys}}^{(i)} P(E_{\text{sys}}^{(i)}) \\ &= Q(\beta, V, N)^{-1} \sum_i E_{\text{sys}}^{(i)} \exp\left(-\beta E_{\text{sys}}^{(i)}\right) \\ &= -\frac{\partial \ln Q}{\partial \beta}. \end{aligned} \tag{2.12}$$

From Thermodynamics, we know that the Helmholtz free energy $A(\beta, V, N)$ is related to the internal energy by:

$$E = \frac{\partial \beta A}{\partial \beta}, \tag{2.13}$$

and by comparing to Equation 2.12 we deduce:

$$\beta A(\beta, V, N) = -\ln Q(\beta, V, N). \tag{2.14}$$

The important feature of Equation 2.14 is that it relates Q, which explicitly depends upon the microscopic details of the system, to the macroscopically measurable quantity A.

2.1.4 The Classical Approximation

By considering states that are eigenstates of the system Hamiltonian, and by postulating that the system is equally likely to be found in any of these states, we have been able to establish relationships between the microscopic quantities and thermodynamic variables (e.g. Equations 2.6 and 2.14). This, however, has been entirely formulated in terms of a quantum mechanical system. Even if we could solve Schrödinger's equation for many-body systems such as bulk liquid water, we would be left with the impossible task of having to compute thermal averages such as:

$$\langle X \rangle = \frac{\sum_i \exp\left(-\beta E_{\text{sys}}^{(i)}\right) \langle i | \hat{X} | i \rangle}{\sum_i \exp\left(-\beta E_{\text{sys}}^{(i)}\right)}. \tag{2.15}$$

Numerical evaluation of all expectation values of the operator $\hat{X}$ of non-negligible weight is infeasible. Fortunately, we are able to carry out a canonical ensemble calculation independently of the exact solution to Schrödinger's equation. This can easily be seen by writing the canonical partition function as:

$$\begin{aligned} Q(\beta) &= \sum_i \mathrm{e}^{-\beta E_{\mathrm{sys}}^{(i)}} = \sum_i \langle i| \, \mathrm{e}^{-\beta \hat{H}} \, |i\rangle \,, \\ &= \mathrm{Tr}\left[\mathrm{e}^{-\beta \hat{H}}\right], \end{aligned} \tag{2.16}$$

where 'Tr [**M**]' denotes the trace of the matrix **M**. Since the trace is independent of the representation of the matrix, once we know $\hat{H}$, we can use any complete set of wavefunctions to compute $Q(\beta)$. For convenience, we can use the set of eigenfunctions of the position or momentum operator. In the classical limit, Equation 2.16 can be written as (see Appendix A):

$$Q(\beta) = \mathrm{Tr}\left[\mathrm{e}^{-\beta \hat{H}}\right] \rightarrow \frac{1}{h^{dN} N!} \int \mathrm{d}\mathbf{r}^N \mathrm{d}\mathbf{p}^N \exp\left(-\beta H(\mathbf{r}^N, \mathbf{p}^N)\right), \tag{2.17}$$

where $H(\mathbf{r}^N, \mathbf{p}^N)$ is the classical Hamiltonian of the system that depends upon the particle positions $\mathbf{r}^N$ and momenta $\mathbf{p}^N$, h is Planck's constant and d is the dimensionality of the system. The factor of $1/N!$ has been included to avoid overcounting indistinguishable microstates. The easiest (but incorrect) way to justify this is that the square of the quantum mechanical wave function is invariant to permutation of identical particles.[2]

The classical Hamiltonian can be written as:

$$H(\mathbf{r}^N, \mathbf{p}^N) = K(\mathbf{p}^N) + U(\mathbf{r}^N) \tag{2.18}$$

where $K(\mathbf{p}^N) = \sum_i^N \mathbf{p}_i^2/2m$ and $U(\mathbf{r}^N)$ are the total kinetic and potential energy of the system, respectively (m is the mass of a particle; for convenience, we are assuming that all N particles are identical). Note that $K(\mathbf{p}^N)$ does not depend upon the particle positions and similarly, $U(\mathbf{r}^N)$ is independent of the particle momenta. Because of this, we can integrate over the momenta to write the classical

[2]There is actually no need to invoke quantum mechanics to justify the inclusion of $1/N!$. In fact, it is needed in entirely classical systems too. See References [28–30] for detailed discussions.

partition function as:

$$Q(\beta) = \frac{1}{\Lambda^{3N} N!} \int \mathrm{d}\mathbf{r}^N \exp\left(-\beta U(\mathbf{r}^N)\right) \tag{2.19}$$

where $\Lambda \equiv h/\sqrt{2\pi m k_\mathrm{B} T}$ is the *thermal de Broglie wavelength*. From Equation 2.19, we see that we are left with the task of performing a Boltzmann-weighted integration over all particle positions. It is therefore convenient to define the *configurational partition function*:

$$Z = V^{-N} \int \mathrm{d}\mathbf{r}^N \exp\left(-\beta U(\mathbf{r}^N)\right) \tag{2.20}$$

Equation 2.20 encompasses essentially all of theoretical and computational chemistry: the biggest challenges in this field are sufficiently sampling configuration space (indicated by $\int \mathrm{d}\mathbf{r}^N$) and defining accurate interaction potentials (indicated by $U(\mathbf{r}^N)$). Figure 2.1 highlights typical methods that are used to define interactions between atoms and molecules, and the balance between computational cost and accuracy[3] that one must consider before starting a molecular simulation. Generally speaking, more affordable techniques are less accurate, but allow for an easier exploration of configuration space. Most of the work in this thesis therefore uses these less expensive methods, although examples of the more sophisticated techniques will be seen. In fact, Chapter 7 is entirely devoted to testing the accuracy of commonly used approximations for $U(\mathbf{r}^N)$ in the case of methane hydrate.

2.1.5 Free Energies Along Order Parameters

In Chapter 1 the concept of 'free energy' was used to explain, in general terms, why an INA enhances ice nucleation. Although we have already encountered the Helmholtz free energy (see Equation 2.14), this is not quite the same as the free energy discussed in Chapter 1. For this, the concept of an *order parameter* needs to be introduced. An order parameter is a function that depends upon the positions of all, or a subset of, the particles and measures the degree of order in

[3]To be clear, 'accuracy' in this context is referring to how closely $U(\mathbf{r}^N)$ approximates the real interaction potential. A less accurate form of $U(\mathbf{r}^N)$ that is also less expensive may provide statistically more meaningful results through a better sampling of configuration space.

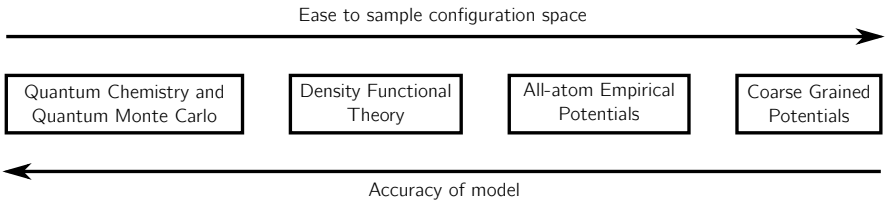

Figure 2.1: Typical methods for computing the interaction potential. Quantum chemistry, quantum Monte Carlo and density functional theory all explicitly treat the quantum nature of the electrons. In the case of all-atom empirical potentials and coarse grained potentials, these electronic degrees of freedom have been integrated out (and the computational cost is significantly reduced). Generally, the accuracy of these methods decreases from left to right, but so too does the computational cost. It is easier to sample configuration space with the less expensive methods.

the system. In the context of crystallisation, an order parameter will typically take the form:

$$Q(\mathbf{r}^N) \approx \begin{cases} 0 & \text{for a liquid configuration,} \\ 1 & \text{for a crystal configuration.} \end{cases} \tag{2.21}$$

A restricted configurational partition function can be defined:

$$Z(Q) = V^{-N} \int \mathrm{d}\mathbf{r}^N \delta\left(Q(\mathbf{r}^N) - Q\right) \exp\left(-\beta U(\mathbf{r}^N)\right). \tag{2.22}$$

where $\delta(x)$ is the Dirac delta function. $Z(Q)$ counts all the microscopic configurations that yield a value Q and weights them by the Boltzmann factor. The restricted configurational partition function is related to the full configurational partition function by an integration over Q:

$$Z = \int \mathrm{d}Q\, Z(Q). \tag{2.23}$$

The ratio of $Z(Q)$ and Z is the normalised equilibrium probability density of observing $Q(\mathbf{r}^N) = Q$:

$$P(Q) = \frac{Z(Q)}{Z} = \langle \delta(Q(\mathbf{r}^N) - Q) \rangle. \tag{2.24}$$

A free energy as a function of Q is now defined:

$$\mathrm{e}^{-\beta A(Q)} = Q_0 Z(Q), \tag{2.25}$$

where Q_0 is the physical unit of the reaction coordinate.[4] It is this free energy that is discussed in Chapter 1. Note that using Equation 2.25, this free energy can also be written as:

$$A(Q) = -k_\mathrm{B} T \ln P(Q) + \text{constant}. \tag{2.26}$$

This is most informative: if we are confident that Q can adequately distinguish between configurations corresponding to the liquid and those corresponding to crystalline ice, then Equation 2.26 states that the 'free energy of the liquid' is dependent upon the equilibrium probability of observing the system in a microstate whose configuration is 'liquid-like' (and similarly for ice). The origin of the barrier separating the liquid and crystal basins (where we can now define a 'basin' as the collection of Q values around a minimum on the free energy surface) is now easily understood as the existence of fewer thermally accessible microstates that give values of Q intermediate between liquid and ice. The role of a good INA is therefore to: increase the number of, or energetically stabilise, these intermediate microstates; or reduce the number of, or energetically destabilise, microstates belonging to the liquid basin.

2.2 Classical Nucleation Theory

2.2.1 Homogeneous Nucleation

In this section, we will briefly discuss *classical nucleation theory* (CNT). To begin, we will consider homogeneous nucleation. In CNT, the metastable liquid phase transforms into the stable crystalline phase through the formation of a small crystal nucleus. This is shown schematically in Figure 2.2. A chemical potential difference $\Delta\mu \equiv \mu_\mathrm{crys} - \mu_\mathrm{liq}$ exists between the liquid and the crystal. Note that as the crystal is more stable than the liquid, $\Delta\mu$ is a negative quantity. The

[4] The inclusion of Q_0 is justified by noting that $Z(Q)$ is a density, whereas $\exp(-\beta A(Q))$ is dimensionless.

crystal nucleus is assumed to be spherical. The change in free energy $\Delta G(R)$ upon forming a nucleus of radius R is:[5]

$$\Delta G(R) = \frac{4\pi}{3} R^3 \rho_{\text{crys}} \Delta\mu + 4\pi R^2 \gamma_{\text{cl}}, \tag{2.27}$$

where γ_{cl} is the crystal-liquid surface free energy and ρ_{crys} is the number density of the bulk crystal. The first term is the gain in free energy obtained by forming the more stable crystalline phase. The second term is positive and is the cost of forming an interface between the liquid and the crystal. For small R, the cost of forming the interface dominates and it is unfavourable to form the crystal nucleus. For larger values of R, the gain from forming the crystal outweighs the cost of forming the interface, and crystallisation proceeds uninhibited. This means that there must be an intermediate value $R = R^*$ where the free energy is maximal. The free energy profile according to CNT is shown in Figure 2.3. It is straightforward to show that:

$$R^* = -\frac{2\gamma_{\text{cl}}}{\rho_{\text{crys}} \Delta\mu}, \tag{2.28}$$

and that:

$$\Delta G(R^*) = \frac{16\pi \gamma_{\text{cl}}^3}{3(\rho_{\text{crys}} \Delta\mu)^2}. \tag{2.29}$$

A nucleus with $R = R^*$ is called a *critical nucleus*. Nuclei smaller than the critical nucleus will tend to 'melt' whereas those larger than the critical nucleus will tend to lead to crystallisation of the entire system.

In the context of Section 2.1.5, R is an order parameter that describes the state of the system: when $R \ll R^*$ we would be comfortable saying that the system is in the liquid state; and for $R \gg R^*$ we would be confident in describing the system as crystalline. Using Equation 2.26 we see that $\Delta G(R^*)$ determines the equilibrium probability of observing a critical nucleus:

$$P(R^*) \propto \exp\left(-\beta \Delta G(R^*)\right) \tag{2.30}$$

[5]We have changed from the symbol 'A' to 'G' for the free energy as the system under consideration now is at constant pressure rather than constant volume. Consequently, the free energy is the 'Gibbs free energy' rather than the 'Helmholtz free energy'. The general concepts introduced in Section 2.1 also apply to the Gibbs free energy.

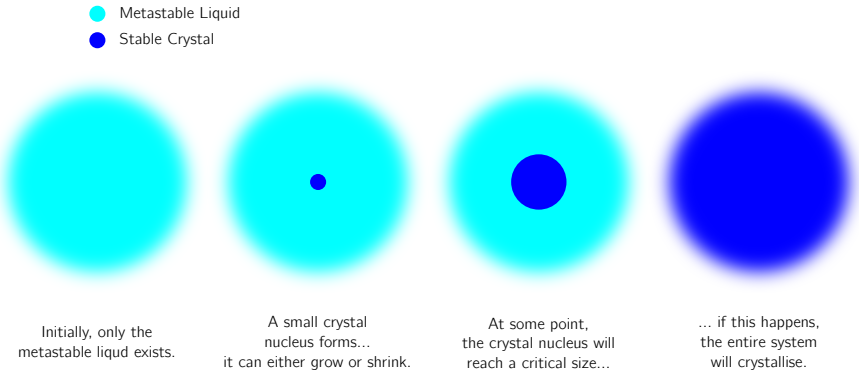

Figure 2.2: Schematic picture of classical nucleation theory (CNT). In CNT, the crystal nucleus is spherical. Furthermore, macroscopic values for e.g. the surface tension and the crystal density are used.

In order to obtain a rate, this probability of reaching the top of the CNT barrier needs to be multiplied by a kinetic prefactor κ. The rate according to CNT is thus:

$$J_{\mathrm{CNT}}(T) = \kappa \exp\left(-\frac{16\pi\gamma_{\mathrm{cl}}^3}{3k_{\mathrm{B}}T(\rho_{\mathrm{crys}}\Delta\mu)^2}\right). \tag{2.31}$$

The kinetic prefactor takes the form [31]:

$$\kappa = Z\rho_{\mathrm{liq}} f_{R^*}, \tag{2.32}$$

where ρ_{liq} is the number density of the metastable liquid, f_{R^*} is the rate at which particles attach to the critical nucleus and Z is the Zeldovich factor:

$$\begin{aligned} Z &= \frac{1}{4\pi\rho_{\mathrm{crys}}R^{*2}}\sqrt{\frac{|\Delta G''(R^*)|}{2\pi k_{\mathrm{B}}T}}, \\ &= \frac{(\rho_{\mathrm{crys}}\Delta\mu)^2}{8\pi\rho_{\mathrm{crys}}\sqrt{k_{\mathrm{B}}T\gamma_{\mathrm{cl}}^3}} \end{aligned} \tag{2.33}$$

where $\Delta G''(R^*)$ is the second derivative of the free energy with respect to R, evaluated at $R = R^*$. The Zeldovich factor approximately corrects for barrier recrossing (i.e. not all nuclei with $R > R^*$ lead to crystallisation).

Although conceptually appealing, CNT has its drawbacks. First of all, it assumes that the most stable phase forms directly from the supercooled liquid and neglects the possibility for a less stable, but more accessible phase to nucleate first (this is known as Ostwald's 'rule of stages'). Another issue with CNT is

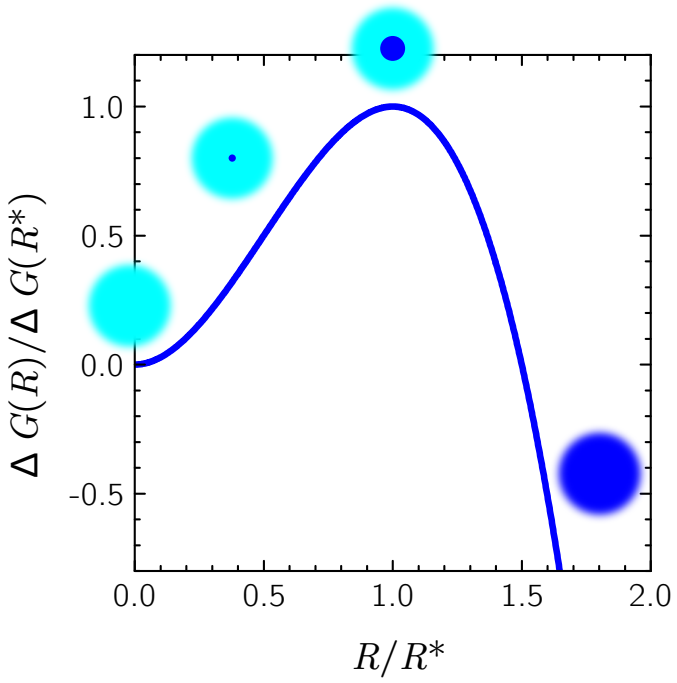

Figure 2.3: Free energy profile according to CNT. For small values of R, the cost of forming the liquid-crystal interface dominates and the free energy increases. For large values of R, however, the gain in forming the more stable crystal outweighs this cost, and crystallisation proceeds uninhibited. The critical nucleus has a radius $R = R^*$: this is the crossover point between these to regimes. The images inset correspond to those shown in Figure 2.2.

that, although we have discussed its formulation in terms of crystallisation, it was originally derived to describe vapour to liquid droplet condensation. CNT therefore lends itself to a picture of the crystal nucleus growing monomer-by-monomer (see e.g. Reference [32] and references therein). The suitability of this assumption will be touched upon in Chapter 6. Other shortcomings of CNT are discussed in Reference [15].

2.2.2 Heterogeneous Nucleation

The extension of CNT to heterogeneous nucleation is straightforward. In fact, the free energy barrier for heterogeneous nucleation can be written as [33]:

$$\Delta G_{\text{het}}(R^*) = f(\theta_{\text{cs}})\Delta G_{\text{hom}}(R^*), \tag{2.34}$$

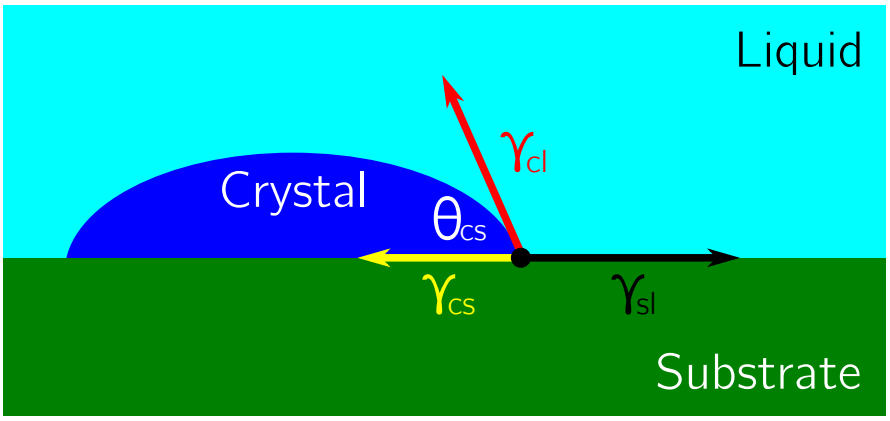

Figure 2.4: Contact angle of an ice nucleus growing on a substrate. At mechanical equilibrium, Young's equation states: $\gamma_{sl} = \gamma_{cs} + \gamma_{cl}\cos\theta_{cs}$.

where $\Delta G_{hom}(R^*)$ is the barrier for homogeneous nucleation given by Equation 2.29. The function $f(\theta_{cs})$ has the form:

$$f(\theta_{cs}) = (2 + \cos\theta_{cs})(1 - \cos\theta_{cs})^2/4, \tag{2.35}$$

where θ_{cs} is the contact angle between the growing crystal nucleus and the substrate,[6] as illustrated in Figure 2.4. The contact angle is given by Young's equation:

$$\theta_{cs} = \cos^{-1}\left(\frac{\gamma_{sl} - \gamma_{cs}}{\gamma_{cl}}\right) \tag{2.36}$$

where alongside the crystal-liquid interfacial free energy γ_{cl}, the substrate-liquid and crystal-substrate interfacial free energies have been introduced (γ_{sl} and γ_{cs}, respectively). Clearly, for $\theta_{cs} = 0°$, nucleation is barrierless in the CNT approximation, whereas if $\theta_{cs} = 180°$, then nucleation occurs homogeneously.

As well as suffering from the same problems as CNT for homogeneous nucleation, another problem with CNT applied to heterogeneous nucleation is that it does not give us any *a priori* indication of whether a material is going to be a good INA. Even using physical intuition, one cannot reasonably guess if a nucleus of ice is going to 'wet' a surface (i.e. $\theta_{cs} < 90°$). It offers no insight into helping us understand which properties of a surface will actually help ice form. What is more, CNT applied to heterogeneous nucleation is not even that useful experimentally, due to the difficulties in measuring θ_{cs}; it is even questionable if

[6] The term 'substrate' is used here to mean the nucleating surface of an INA.

θ_{cs} is a well-defined quantity.

2.3 The 'Requirements' for an Ice Nucleating Agent

Given the importance of ice nucleation, attempts have of course been made to try and understand which properties of a substance affect its ice nucleating ability. One particularly famous set of 'requirements' for a good INA are those of Pruppacher and Klett [1]. These can be paraphrased as:

1. **Insolubility Requirement**. In general, INAs are highly water-insoluble. The obvious disadvantage of a soluble substrate is that its tendency to disintegrate under the action of water prevents it from providing a rigid substrate upon which an ice nucleus can grow. Furthermore, the presence of salt ions lowers the effective freezing temperature.
2. **Size Requirement**. Aerosol particles in the Aitken size range (radii less than 0.1 μm) are less efficient INAs than 'large' aerosol particles. It is generally said that good INAs must be on the size of, or larger than, the critical nucleus. A dependence of aerosol particle chemistry cannot be entirely ruled out, however, as good INA such as silicate particles are mostly confined to the 'large' size range.
3. **Chemical Bond Requirement.** Considering the fact that an ice crystal is held together by hydrogen bonds, it is reasonable to assume that a good INA must have similar hydrogen bonds available at its surface.[7]
4. **Crystallographic Requirement.** The geometrical arrangement of bonds at the substrate surface is often of equal or greater importance than their chemical nature. The closer the crystallographic match to ice, the better the ice nucleating ability of a material will be.
5. **Active Site Requirement.** Heterogeneous ice nucleation is a very localised phenomenon in that it proceeds at distinct active sites on a substrate's surface. Sites at which ice nucleation is initiated are also sites that are capable of adsorbing water molecules. One may distinguish between three types of active site: (i) morphological surface inhomogeneities such as steps, cracks or a cavity at the surface of the INA; (ii) chemical inhomogeneities in the surface, generally caused by the presence of a foreign ion that is

[7]The 'chemical bond requirement' can also be interpreted as a 'hydrophilicity' requirement.

> hydrophilic relative to the rest of the solid surface; and (iii) electrical inhomogeneities other than ions, such as sharply defined boundaries between surface regions of different electric field sign.

There are many exceptions to these requirements. For example, crystalline soluble salts such as ammonium sulphate have been observed to nucleate ice in saturated solution droplets [34]. This has led some to propose that this should be rephrased as a 'solid' requirement [7], although this would preclude ice nucleation by surfactant molecules [35]. Amorphous materials such as glassy atmospheric aerosols [36], organic aerosol [37] and soot [8–10] can facilitate ice formation and these bear little resemblance to the crystal structure of ice. These criteria have neither served as a full set of guidelines to identify good INAs, nor have they aided the systematic improvement of ice nucleation inhibitors or promoters. When compared to the field of chemical catalysis [22], along with other areas in materials science [38], the ice nucleation community falls short.

Of the 'requirements' outlined above, it is the last three that lend themselves most naturally to study by molecular simulation. In particular, the 'chemical bond requirement' and the 'crystallographic requirement' will be investigated in detail at various points throughout the thesis. Indeed, the 'crystallographic requirement' is the main focus of the next chapter.

Chapter 3

The Role of Lattice Mismatch

3.1 Grand Canonical Monte Carlo

The *Monte Carlo method* (MC), in principle, allows efficient sampling of the canonical ensemble. When using rigid molecules, this is achieved by translating or rotating a molecule at random and accepting the move with probability:

$$P^{\text{acc}}_{\text{trans,rot}} = \min\left(1, \exp\left[-\beta\left(U_{\text{n}}(\mathbf{r}'^{N}) - U_{\text{o}}(\mathbf{r}^{N})\right)\right]\right), \tag{3.1}$$

where the labels 'o' and 'n' denote the configurations before and after the trial move respectively.

If the problem of interest is adsorption, a far more efficient sampling method is grand canonical Monte Carlo (GCMC), where instead of fixing N, the chemical potential μ is constant. Thus, along with the translational and rotational moves, attempts to insert or delete a molecule are also performed (for insertion, a molecule is given a random position and orientation; for deletion, a molecule is chosen at random). The probability of accepting an insertion or deletion is:

$$P^{\text{acc}}_{\text{insert}} = \min\left(1, \frac{zV}{N+1}\exp\left[-\beta\left(U_{\text{n}}(\mathbf{r}'^{N+1}) - U_{\text{o}}(\mathbf{r}^{N})\right)\right]\right),$$

$$P^{\text{acc}}_{\text{delete}} = \min\left(1, \frac{N}{zV}\exp\left[-\beta\left(U_{\text{n}}(\mathbf{r}'^{N-1}) - U_{\text{o}}(\mathbf{r}^{N})\right)\right]\right), \tag{3.2}$$

where z is the absolute activity:

$$z = \frac{\exp(\beta\mu)}{\Lambda^3}. \tag{3.3}$$

The reason why GCMC is more suited to adsorption problems is that the vapour phase is not explicitly sampled meaning that the size of the simulation cell is smaller compared to what would be required in a canonical MC simulation and that moves are not wasted on 'uninteresting' rearrangements of the vapour.

The above is only meant to act as a very brief outline of MC and GCMC, and is far from a complete overview of the various Monte Carlo methods available. For more details and the derivation of the acceptance probabilities given in Equations 3.1 and 3.2, the reader is referred to References [25] and [39].

3.2 Introduction

As we saw in the previous chapter, there are numerous textbook arguments (e.g. insolubility, size or defects) for a substrate to be an effective INA but an important one, which led to AgI being used as a cloud seeding agent [40], is a good crystallographic match with bulk hexagonal ice [1, 33]. The crystallographic match can be quantified by the disregistry (or "mismatch"), defined in a simplified manner as:

$$\delta \equiv \frac{a_{0,\mathrm{S}} - a_{0,\mathrm{i}}}{a_{0,\mathrm{i}}}, \tag{3.4}$$

where $a_{0,\mathrm{S}}$ is the lattice constant of the substrate and $a_{0,\mathrm{i}}$ is the bulk lattice constant of ice $\mathrm{I_h}$. The original theory, developed within the context of classical nucleation theory by Turnball and Vonnegut [33], envisages a situation where the growing crystal can be pictured as consisting of regions of good fit (strained slightly to fit the underlying lattice) bounded by line dislocations.[1] Furthermore, through the interpretation of early experimental data (in particular from low energy electron diffraction experiments), the concept of an ice-like bilayer forming at pristine metal surfaces was developed in which the water molecules occupy 2/3

[1] In the original theory, Turnball and Vonnegut define two regimes; a low disregistry regime ($\delta < 0.2$), where the crystal grows coherently with the substrate and a high disregistry regime, where the crystal grows incoherently with the substrate. All results presented here fall into the low disregistry regime. It should be noted, however, that the differences from the original theory found in this work give this decomposition into two regimes little meaning.

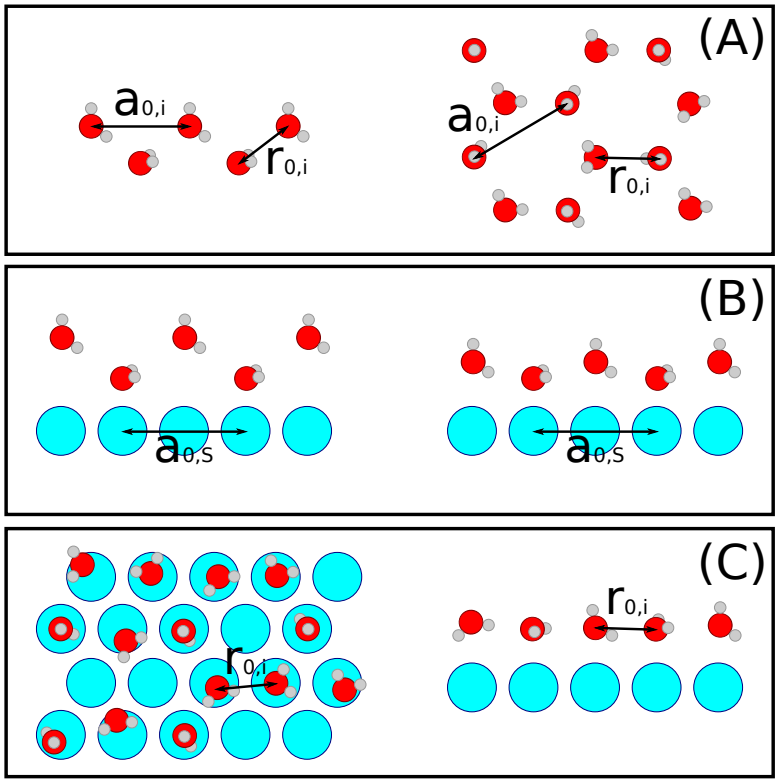

Figure 3.1: Schematic of the first water overlayer at a surface. (A) Part of a single isolated ice bilayer shown from the side (left) and top (right). The bulk ice lattice constant $a_{0,\mathrm{i}}$ and nearest neighbour oxygen–oxygen distance $r_{0,\mathrm{i}}$ are labelled. (B) Ice bilayer at the surface according to the traditional theory. On the left, the ice bilayer has strained to fit the lattice of the surface, which has a lattice constant $a_{0,\mathrm{S}}$ slightly larger than bulk ice. On the right, the bilayer is allowed to relax normal to the surface. (C) Along with relaxation normal to the surface, one can also envisage a scenario where changes in morphology in the plane of the surface are favourable.

of adsorption sites (2/3 monolayer coverage) [41].

This theory is often used as an explanation for the excellent ice nucleating abilities for many materials, including kaolinite [1], AgI [33], and long chain aliphatic alcohols [35]. However, in certain specific cases, it has been questioned what role, if any, the lattice mismatch plays in ice nucleation [42–44]. Indeed, DFT calculations on idealised ice bilayers at different metal surfaces [45] have shown that through relaxation normal to the surface, the water molecules manage to maintain near constant hydrogen bonding energy, in contradiction to the traditional theory (Figure 3.1.B). Also, recent experimental evidence, especially from scanning tunnelling microscopy, has revealed a number of complex structures of water on the first overlayer of metal surfaces, inconsistent with the bilayer model [46–49].

Here, using GCMC and simple intermolecular potentials, the effect of the lattice constant of the substrate on the structure of the first water overlayer will be probed. It should be noted that of the four basic modes of heterogeneous ice nucleation [1], it is the deposition mode that is being most closely simulated, in which water is adsorbed directly from the vapour phase onto the surface where it forms the ice phase. Although there have been previous theoretical works that have used model surfaces to study the structure of interfacial water (see e.g. References [50–53]), none have explicitly tackled the effect of changing the lattice parameter of the substrate alone on the structures that form in the first water overlayer. What we will see is that, in contrast to the traditional theory, the first water overlayer consists of a significant fraction of non-hexagonal arrangements, with smaller lattice constants favouring smaller sized rings and fully hexagonal overlayers only observed on substrates with expanded lattice constants. As most molecules in the contact layer are of a similar height, it will be argued that the nearest neighbour oxygen–oxygen distance $r_{0,\mathrm{i}}$ becomes more relevant than the bulk ice lattice constant $a_{0,\mathrm{i}}$ in describing the structures observed (see Figure 3.1), consistent with recent experimental observations on close packed metal surfaces [46, 54].

A qualitative overview of the influence of the lattice constant alone on the structures that form in the first water overlayer is desirable. DFT, whilst proven to be a useful method for investigating interfacial water (see e.g. References [46, 49, 54–56]), would be inappropriate for such a study not only because of the restriction in system size that could be used, but also because it would not be possible to freely vary the lattice parameter of the surface without changing other properties of the surface. To this end, interaction potentials will be used to model the interaction between the water and the surface. Even though the use of simple potentials cannot describe the 'chemistry' present at real surfaces, this can be advantageous in that the lattice parameter can be effectively decoupled from other properties of the surface. However, to be as careful as possible to ensure the observed trends are general and not just applicable to one type of model, two types of water–surface potential and two standard water models have been used. Details of these tests can be found in Appendix B. It must be stressed that

this work is not aiming to model water layers at real surfaces, but to probe an often-used, long-standing theory that only concerns the geometrical distribution of adsorption sites, which has so far proved inadequate in a number of cases [46–49].

3.3 Models and Computational Setup

The first type of surface employed consists of four layers of points each in an hexagonally closed packed arrangement, interacting with the oxygen atoms of the water molecules through a Lennard-Jones interaction. The adsorption sites are located at the HCP and FCC sites of this surface, giving rise to a honeycomb arrangement of potential energy minima for the water monomer – this surface shall be referred to as the "honeycomb" surface. The second surface used is an explicitly defined external potential acting on the oxygen atoms of the water molecules. The adsorption sites on this surface are in an hexagonally close packed arrangement – this surface is referred to as the "HCP" surface. On both surfaces, the monomer adsorption energy was chosen to be comparable to, but slightly stronger than, the TIP4P dimer binding energy ($-26.09\,\mathrm{kJ/mol}$) [57] and similar to the optB88-vdW functional [58, 59] value on Ag(111) ($-27.112\,\mathrm{kJ/mol}$) [60].[2] Full details of both these surfaces are given in Appendix B.

The simulations presented here all used the TIP4P potential, a rigid simple point charge model, to represent the water–water interactions [61]. The TIP4P water model has been shown to reproduce the phase diagram of water qualitatively well [62], even though it predicts a melting temperature of ice $\mathrm{I_h}$ to be $T_\mathrm{m} = 232\,\mathrm{K}$. As mentioned previously, in order to see how dependent the results are on the choice of water force field used, the simulations have also been performed with the SPC/E water model [63], with good agreement between the two models obtained (see Appendix B).

GCMC simulations were performed on six different slab geometries for the two types of surfaces at 215 K corresponding to a supercooling of 17 K (the SPC/E calculations were performed at 198 K to achieve the same level of supercooling),

[2] Although the values used for the binding and barrier height have been taken from DFT using the optB88-vdW functional, this is only to give a binding energy that is comparable to, but stronger than, the TIP4P-dimer binding energy and a reasonable barrier height. It is not expected, nor indeed has any attempt been made, for the simulations to mimic water on Ag(111).

each with a different lattice constant, using the simulation package DL_MONTE [64]. The surface areas varied between slabs (ranging from $47.51 \times 41.14\,\text{Å}^2$ to $55.41 \times 47.99\,\text{Å}^2$) but the cell volume was kept constant between simulations, ensuring there was a vacuum of at least 42.65 Å between periodic images in the z-direction. Full periodic boundary conditions were employed and electrostatic interactions were treated using the three-dimensional Ewald method [39] and short range interactions were calculated by employing a spherical cut-off of 17 Å. A GCMC step could be either a molecule translation, rotation, insertion or deletion and these were attempted with a ratio of 33:33:17:17. The surface was fixed throughout. Simulations were run until a single overlayer of water was obtained. After the GCMC simulations, final structures were annealed at 15 K and water molecules whose oxygen atom was greater than 3.5 Å above the surface were removed. The structures were then annealed for a further 5×10^6 MC steps at 15 K.

3.4 Results

GCMC is a powerful technique most often used as a method for calculating adsorption isotherms. However, here it is being used to efficiently sample configuration space when generating the single water overlayers.[3] The GCMC simulations generated an array of structures in the first overlayer on the different surfaces, with coverages on the HCP surfaces ($\delta = -0.07$ to 0.10) increasing approximately linearly from ∼0.61 monolayer coverage to ∼0.78 monolayer coverage. This is in contrast to the traditional theory in which 2/3 monolayer coverage is predicted. Figure 3.2 presents snapshots of water layers obtained for selected values of δ on both the honeycomb and HCP surfaces. One trend is immediately clear; as the mismatch is increased, more open ring-like structures are favoured, whereas as δ is decreased, more closely packed arrangements start to be seen. However, perhaps the most striking feature of these pictures is that on the $\delta = 0.00$ surfaces, there

[3] Although GCMC provides an efficient manner by which to explore configuration space, there is no guarantee that it will generate the lowest energy structures. In fact, it should also be noted that the periodicity imposed by the boundary conditions may play a role in preventing the lowest energy structures being found. However, by using large simulation cells (between 288 and 400 adsorption sites), this finite size effect is being reduced as much as possible within the available computational resources.

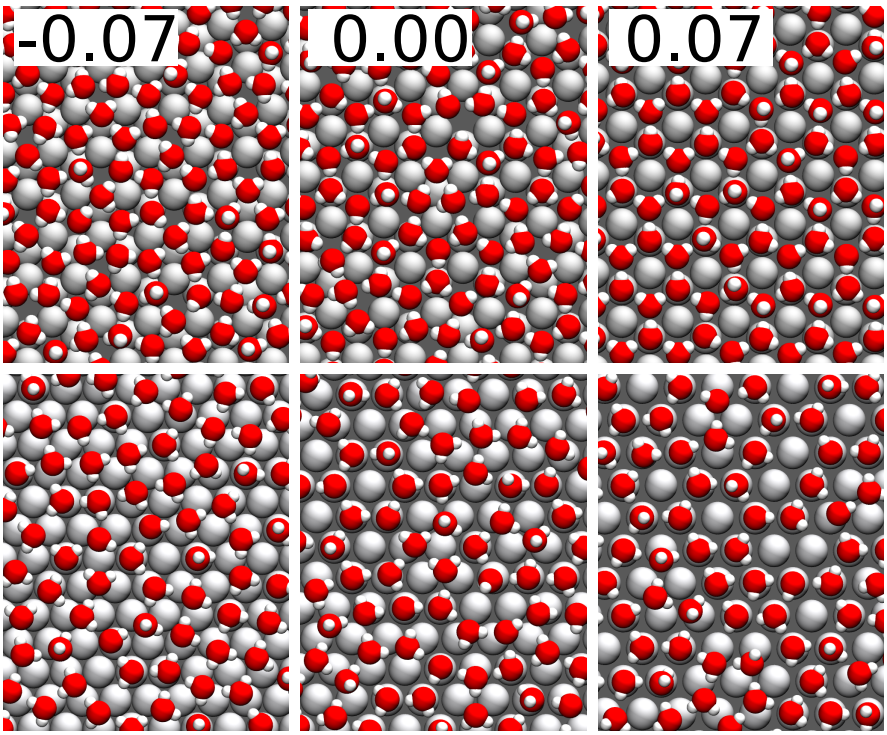

Figure 3.2: Snapshots for selected values of δ. The top row shows the honeycomb surface and the bottom row shows the HCP surface. The value of δ (disregistry, as defined in Eq. 3.4) is shown in the top left corner. It can be seen that on both $\delta = 0.00$ surfaces there is a significant proportion of non-hexagonal arrangements of water molecules, the effect being more pronounced on the HCP surface. As δ is increased hexagons become more favoured and conversely, as δ is decreased, an increase of more compact structures is seen. For clarity, only a section of the entire simulation cell is shown and the atoms have been depicted to be the same size.

is still a significant proportion of non-hexagonal structures present, with the only fully hexagonal overlayers being formed on the honeycomb surface for $\delta \geq 0.07$.

Presented with these results, it is interesting to ask why such an array of structures is seen in the different overlayers and why it is only on larger δ that hexagons start to become more favourable. To address these questions, it is informative to look at the radial distribution functions (RDFs) and the energies of different overlayers, which are given in Figure 3.3 (the corresponding plots for the honeycomb surface are included in Appendix B). From the RDFs, it can be seen that the position of the first peak is quite insensitive with regard to the underlying substrate, lying at approximately 2.75 Å. The significance of this can be understood by revisiting the definition of the disregistry, δ (Equation 3.4) and looking more closely at the 'ice bilayer' model.

In bulk ice I_h, the oxygen atoms lie in layers perpendicular to the c-axis,

with each layer being built up of puckered hexagonal rings (a single ice 'bilayer'), arranged in ABABAB... stacks. A schematic of a single ice bilayer is shown on the left of Figure 3.1.A. The lattice constant, $a_{0,\mathrm{i}}$ corresponds to the distance between two nearest coplanar oxygen atoms and not the distance between two nearest hydrogen bonded water molecules (labelled $r_{0,\mathrm{i}}$ in Figure 3.1). From the RDFs it appears that the water is able to relinquish long range order in favour of maintaining a constant nearest neighbour distance. It should be stressed that this "constant" nearest neighbour distance refers to the position of the first peak and that this peak exhibits a finite width (ranging from just under 2.6 Å to approximately 3.2 Å). In addition, rather than forming an ice bilayer at the surface, a flat monolayer forms (see Figure B.5). With these last two points in mind, simple geometry is enough to explain why such a significant proportion of pentagons is seen at the $\delta = 0.00$ surface; the distance between adjacent sites of potential energy minima $r_{0,\mathrm{S}}$ is not large enough to accommodate a fully hexagonal overlayer with a favourable $r_{0,\mathrm{i}}$. As the surface is stretched, $r_{0,\mathrm{S}}$ becomes larger allowing easier formation of hexagons at the surface. When the surface is compressed, however, one way that the water molecules can arrange themselves such that they keep a favourable $r_{0,\mathrm{i}}$ and bind near the adsorption sites is to adopt structures with smaller O–O–O angles.

The energy profiles in Figure 3.3 are useful in explaining these observations further, where the different contributions to the total energy have been separated (water–water and water–surface interactions)[4] for the different overlayers. The first thing to note is that there is surprisingly little variation in the total, water–water and water–surface interactions as the lattice constant of the substrate is changed (the scale on each panel is $4\,(\mathrm{kJ/mol})/\mathrm{H_2O} \approx 41\,\mathrm{meV/H_2O}$). There is a slight stabilisation in the total energy on the expanded substrates, due to an increase in binding to the surface. More importantly, there is very little change in water–water energy, indicating that water is able to relinquish long range order without incurring a severe energy penalty. This tells us that on surfaces where the

[4]In general, such energy decompositions are arbitrary since it is assumed that the hydrogen bonding within the overlayer does not change upon approaching the surface. When using forcefields, this assumption is correct but in general this is not the case. For a more detailed discussion on this type of energy decomposition see e.g. Reference [45].

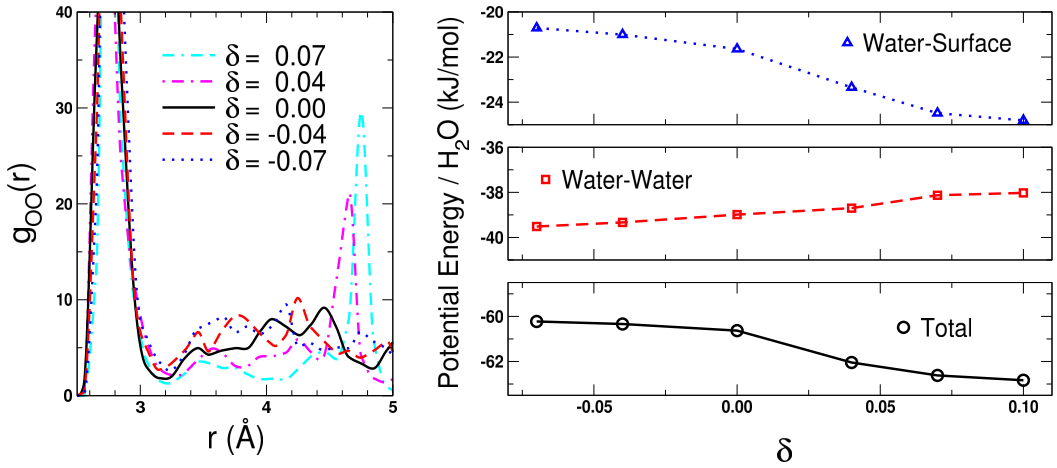

Figure 3.3: Left: radial distribution functions for different δ (HCP surface). Note that the position of the nearest–neighbour peak is essentially invariant as the substrate lattice constant is changed, located at approximately 2.75 Å. Right: contributions to the total energy as δ varies. There is a slight stabilisation in the total energy (bottom) as δ is increased, which arises from an increased binding to the surface (top). This trend is seen as $r_{0,\mathrm{S}}$ is closer to $r_{0,\mathrm{i}}$ for expanded δ, meaning more water molecules can bind close to preferred adsorption sites with a favourable nearest–neighbour distance. The water–water interaction (middle) is very flat, indicating that varying the O–O–O angle is a facile process.

water–water and water–surface interactions are comparable, the cost of forming an overlayer that is commensurate with the substrate (with $r_{0,\mathrm{S}} < r_{0,\mathrm{i}}$) outweighs the gain in binding to the surface. Therefore, provided the water molecules can still bind near to the favoured adsorption sites, the water molecules will alter their O–O–O angles in order to maintain a favourable nearest–neighbour distance.

With these results, one is now in a position to compare to the traditional theory of heterogeneous nucleation [33], where it is assumed that for δ sufficiently small, the growing crystal will strain to fit the substrate lattice. What is unaccounted for in this formalism (and what is observed in these simulations), however, is the ability of the first overlayer to change topology; firstly, the water molecules bind as a flat layer to increase their binding to the surface and secondly, they distort their O–O–O angles so as to maximise their nearest–neighbour water–water interactions. To make a semi-quantitative comparison to the traditional theory, one can try to estimate the stability of the traditional ice bilayers relative to the structures found through GCMC. In order to do this, a single ice bilayer (consisting of 24 molecules) was taken from a bulk ice $\mathrm{I_h}$ structure and geometry optimised for various sizes of the lattice constant using the `GROMACS` simulation package [65].

Two types of geometry optimisations were performed; one set where the positions of the water molecules were constrained such that the c/a ratio is constant and another set where the heights of the water molecules were allowed to vary, only constraining the layer to maintain hexagonal symmetry.[5] The results from this analysis are shown in Figure 3.4, where one can see that the structures generated by GCMC are more stable by approximately 2-3 kJ/mol per H_2O relative to the most stable bilayer (the total adsorption energy is shown in the bottom panel). From inspection of the contribution to the total energy arising from water–water interactions (middle panel) it can be seen that, on the whole, the hydrogen bonding network is slightly destabilised compared to the bilayer in which the height of the water molecules is allowed to relax. In fact, for $\delta < 0.03$, the conventional ice bilayer has a more stable hydrogen bonding network than the GCMC overlayers. Where the GCMC overlayers gain their overall stabilisation, however, is by binding in a flat layer to the surface; the increase in water–surface interaction offsets the very small energy penalty in forming a flat hydrogen bonded network relative to a buckled one.

3.5 Conclusions

In this chapter, the role of lattice mismatch alone on the structure of the first water overlayer has been investigated through the use of GCMC and model hexagonal surfaces. It has been found that, in direct constrast to the traditional theory, the first overlayer alters its topology away from that of the traditional bilayer model in order to maximise its interaction with the surface and in doing so, suffers little energy cost regarding its water–water interactions. This observation has been explained by the fact that once the water molecules bind at equal heights above the surface, there is very little energy associated with changing the O–O–O angles and that independent of the underlying substrate, the water molecules maintain a near constant O–O nearest neighbour distance. The results also suggest that, unless there is an inherent corrugation of the adsorption sites on the surface, one

[5]The total adsorption energy of the bilayers at the HCP surface by assuming that all water molecules lie directly above the adsorption sites and that the low-lying water molecules are at the optimal height. As one knows the functional form of the water surface interaction, calculating the energy of the high-lying water molecules is trivial.

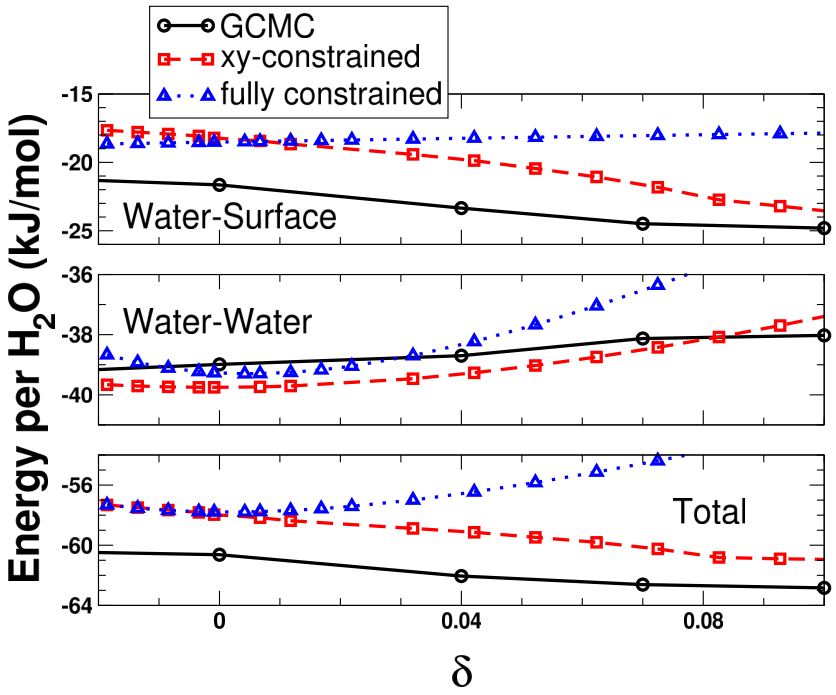

Figure 3.4: Comparison to the traditional theory of heterogeneous ice nucleation. The bottom panel shows the total energy of the different types of overlayer at the HCP surface, the middle panel shows the water–water energy and the top panel shows the interaction energy of the water with the surface. "Fully constrained" is a bilayer whose c/a ratio was kept constant, whereas "xy-constrained" is a bilayer in which the relative heights of the water molecules was allowed to change whilst keeping the lateral coordinates of the water molecules fixed. We can see that by all water molecules binding closely to the surface, the structures generated by GCMC are stabilised relative to the bilayer models.

should not expect to observe fully hexagonal overlayers for substrates that have a mismatch close to zero, consistent with recent experimental findings [54]. Not only does this work confirm the findings that the lattice match alone is insufficient to make a material an efficient INA [42, 43, 50, 66, 67], but it also suggests that a good lattice match in the conventional sense (Equation 3.4) may possibly be detrimental to ice nucleation, as large molecular rearrangement in the first water overlayer may be required to form an ice nucleus. Indeed, if the first water overlayer bears little resemblance to an ice bilayer, it is not entirely obvious what role it plays in heterogeneous ice nucleation; it will be seen in Chapter 5 that the importance of the first layer structure is dependent upon the topography of the substrate.

Chapter 4

Direct Simulation of Ice Nucleation on Kaolinite

The simulations of homogeneous ice nucleation and the DFT adsorption energy calculations presented in this chapter were performed by Zamaan Raza at University College London.

4.1 Molecular Dynamics Simulation

An option other than MC for sampling phase space is *molecular dynamics simulation* (MD). This approach differs from MC in that, rather than attempting to alter the system with trial moves, one numerically solves Newton's equations of motion for the system of interest. For simplicity, only fully classical molecular dynamics will be considered in this brief introduction to MD. This means that not only is the nuclear motion treated within the classical approximation, but also that the intermolecular interactions can be treated analytically without solving the electronic Scrödinger equation. Furthermore, in this brief outline, the pairwise additive approximation will be made such that the total potential energy function can be written as:

$$U(\mathbf{r}^{\mathrm{N}}) = \sum_{i}^{N-1} \sum_{j>i}^{N} u(r_{ij}), \tag{4.1}$$

where r_{ij} is the distance between particles i and j, and $u(r)$ is the potential energy function describing the interaction between them. With a potential energy

function defined, one needs to be able to integrate Newton's equations of motion in order to obtain ensemble averages of desired quantities. One method is the Verlet algorithm, which will now be briefly outlined. The first step is to take a Taylor expansion of a particle's coordinate around time t:

$$r_i(t+\delta t) = r_i(t) + v_i(t)\delta t + \frac{a_i(t)\delta t^2}{2} + \frac{\dddot{r}_i(t)\delta t^3}{6} + \mathcal{O}(\delta t^4); \tag{4.2}$$

and similarly:

$$r_i(t-\delta t) = r_i(t) - v_i(t)\delta t + \frac{a_i(t)\delta t^2}{2} - \frac{\dddot{r}_i(t)\delta t^3}{6} + \mathcal{O}(\delta t^4), \tag{4.3}$$

where v_i is the velocity of particle i, a_i its acceleration and δt is a small change in time ($\dddot{r}_i(t)$ indicates the third derivative of $r_i(t)$ with respect to time). Summing Equations 4.2 and 4.3 we obtain:

$$\begin{aligned} r_i(t+\delta t) &= 2r_i(t) - r_i(t-\delta t) + a_i(t)\delta t^2 + \mathcal{O}(\delta t^4), \\ &= 2r_i(t) - r_i(t-\delta t) + \frac{f_i(t)\delta t^2}{m_i} + \mathcal{O}(\delta t^4), \end{aligned} \tag{4.4}$$

where in the last line Newton's Second Law has been used to incorporate the force on particle i, f_i, where m_i is its mass. An important point to note here is that the first and third order terms cancel out, making the Verlet algorithm more accurate than a simple Taylor expansion. Other integration algorithms do exist (e.g. velocity-Verlet, leapfrog[1] and predictor-corrector algorithms) but the main essence of MD simulation is captured in the Verlet algorithm. In practice, one samples discrete points in time along a trajectory, meaning that a decision about the size of time step δt is required. This involves a trade-off between accuracy and efficiency: as $\delta t \to 0$, Equation 4.4 becomes exact whereas for larger time steps, it will diverge from the true trajectory. Suitability about the choice of time step can be tested by performing a microcanonical simulation of the system and measuring the drift in the total energy. A full overview of molecular dynamics

[1]The velocity-Verlet and leapfrog algorithms generate identical trajectories to the Verlet scheme.

simulation is beyond the scope of this section. For further information, especially regarding the use of thermostats and barostats, the reader is again referred to References [25] and [39], and also Reference [27].

4.2 Introduction

Although the results in the previous chapter provide useful information for trying to understand the suitability of a long standing theory, they do not give any insight into the molecular mechanisms that actually govern heterogeneous ice nucleation. In this chapter, heterogeneous ice nucleation in the presence of a more realistic model surface will be directly probed. Despite a number of computer simulation studies of homogeneous nucleation [68–76], there have been very few that have directly probed heterogeneous nucleation. Yan and Patey [77, 78] have performed an excellent set of MD simulations aimed at investigating the effect of strong electric fields on ice nucleation, finding that ferroelectric cubic ice forms in the region exposed to the electric field. Although this provides some insight into the role of electric fields on nucleation, the fields used are relatively smooth, whereas those exerted by real surfaces are likely to greatly vary on molecular length scales. Solveyra *et al.* [79] have also looked at the effect of confinement on ice nucleation in both hydrophilic and hydrophobic nanopores, using the single-site mW water model [80]. Use of such coarse grained force fields to describe the molecular interactions has the distinct advantage of being able to simulate large length- and time-scales at reasonable computational cost, but would unfortunately be inappropriate for this study, where the electrostatic interactions between the surface and water are significant. The work presented in this chapter is unique in that it is the first to directly simulate the dynamical process of heterogeneous nucleation where the atomic structure of both water and the substrate is taken into account.

As with homogeneous nucleation, there are many computational techniques at our disposal for looking at heterogeneous nucleation. One possible route is to use a free-energy based method such as metadynamics [81, 82] or umbrella sampling [83] (for applications of these methods to homogeneous ice nucleation see References [74–76]). The advantage of methods such as these is that one is able

to obtain free energy barriers to nucleation along a specified reaction coordinate, but with the drawback that the system has to be driven along a predetermined set of collective variables, with no guarantee that the 'true' reaction pathway is being sampled. Another approach is to perform a number of unbiased MD simulations, starting with water in the supercooled liquid state, over suitably long time-scales until the nucleation event is observed. Although adopting such an approach may be seen as computationally inefficient, with recent advances in computer technology and software, the timescales involved are realisable at a reasonable computational cost for small to medium system sizes. Furthermore, by only performing unbiased MD simulations, one is no longer imposing *a priori* the reaction coordinate that the system must traverse. This direct approach has been used to seemingly good effect to study homogeneous ice nucleation, first by Matsumoto *et al.* [68] and subsequently by Jungwirth and co-workers [69–71].

With the aim of understanding heterogeneous ice nucleation, the clay mineral kaolinite has been used as the model ice nucleating agent. Each year, as much as 3000 Tg of mineral dust (naturally occurring crystalline solid compounds) is transported into the troposphere from desert regions [84] where it catalyses the formation of ice [1, 7]. The composition of mineral dust is diverse with quartz, feldspar, calcite and clays all present in significant proportions in typical atmospheric dust samples. Clays are the most frequently observed group in atmospheric mineral dust, of which kaolinite forms a substantial fraction [7]. Apart from being a known effective ice nucleating agent [85–87], the binding of water to the pristine hydroxyl-terminated (001) face has been well characterised theoretically [42, 66], which aids in the analysis of the nucleation simulations.

Kaolinite is a layered silicate mineral with chemical composition $Al_2Si_2O_5(OH)_4$. Each layer consists of a tetrahedral silica sheet alternating with an octahedral alumina sheet, terminated with hydroxyl groups (see Figure 4.1). In the bulk, these layers are bound by hydrogen bonds between the hydroxyl-terminated face and the silica-terminated face, giving rise to facile cleavage along the (001) plane, exposing the hydroxyl- and silicate-terminated faces. It is believed that the hydrophilic hydroxyl-terminated face is the origin of the ice nucleating efficacy of kaolinite, with the textbook explanation being that the pseudo-hexagonal ar-

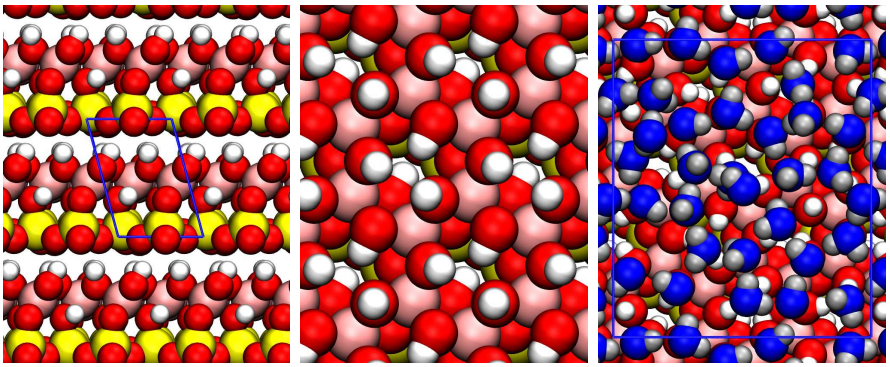

Figure 4.1: Structure of kaolinite. On the left the layered bulk structure of kaolinite is shown. As the layers are bound by hydrogen bonds between the hydroxyl-terminated and silicate-terminated faces, facile cleavage is observed along in the (001) plane. The middle panel shows the hydroxyl-terminated (001) face. DFT calculations [42, 66] show that upon cleavage, 1/3 of OH groups rotate into the plane of the surface, making it amphoteric i.e. able to both accept and donate hydrogen bonds with water. On the right is a snapshot from one of the MD simulations showing the first contact layer of supercooled water. We can see that the water molecules are densely packed and disordered. The colour scheme is: Si, yellow; Al, pink; O, red; and H, white. Water molecules in the first contact layer are shown in blue.

rangement of –OH groups acts as a template upon which the basal face of ice I_h can grow [1]. Despite its attractive simplicity, the validity of this explanation has been questioned; a series of DFT calculations by Hu and Michaelides [42, 66, 88] indicate that the most stable ice-like bilayer at the kaolinite surface is actually hydrophobic with respect to growth of further layers of ice, a property attributed to the amphoteric nature of the hydroxyl-terminated surface; whilst grand canonical Monte Carlo simulations by Croteau *et al.* [43, 89, 90] have shown that only small regions of hexagonal motifs form in the first water overlayer and that these are somewhat stretched relative to bulk ice. In this chapter, the ice nucleation mechanism at the kaolinite (001) surface will be probed directly using MD simulations, in a bid to shed further light onto the process of ice formation in the presence of this important mineral, as well as make inroads into understanding heterogeneous ice nucleation in general.

In what follows, we will see that, rather than the basal face, the prism face grows exclusively from the kaolinite surface. It will be shown that density fluctuations in the supercooled water away from the kaolinite slab play an important role in the heterogeneous nucleation mechanism. The role finite size effects play in the simulations will also be described, before concluding and discussing the

implications of the findings for the macroscopic crystal structure.

4.3 Methods

To model water on kaolinite, the TIP4P/2005 water potential [91] and the CLAYFF potential of Cygan *et al.* [92] have been used. TIP4P/2005, a rigid point charge water model, has been shown to replicate the phase diagram of water qualitatively well along with the transport properties of bulk water, even though it predicts the melting point of ice I_h to be *ca.* 252 K. It also reproduces the experimental bulk densities of liquid water, hexagonal and cubic ice very well, making it a suitable choice for modelling ice nucleation. The CLAYFF potential has been widely used for studying water at various clay mineral interfaces [93–96] and in particular for the study of ice nucleation at kaolinite by grand canonical Monte Carlo [43, 89, 90]. In this approach, the clay atoms are treated as simple point charges with Lennard-Jones interactions, with the only explicit bonding term occurring between the oxygen and hydrogen of the hydroxyl groups. Such flexibility in the model allows CLAYFF to describe a number of different clay structures and phases satisfactorily, as well as the swelling of clays with increased water content [92]. The water-clay interaction was calculated using the standard Lorentz-Berthelot mixing rules [97, 98].

For the MD simulations, a protocol similar to that used by Jungwirth and co-workers [69–71] was followed, who have had much success in direct simulation of homogeneous ice nucleation. To create the homogeneous systems, 192 water molecules were placed in an orthogonal simulation cell with lateral (xy) dimensions[2] of *ca.* 13.2×15.6 Å^2. Due to the small x- and y-dimensions, a small cutoff of 6.5 Å was employed. Electrostatic interactions were calculated using the smooth particle mesh Ewald method, with a pseudo 2D correction [100] for the slab geometry, giving an effective z-dimension of at least 100 Å. The geometry of this system can thus be best described as an infinite slab with two liquid-vapour interfaces. For the heterogeneous system, the kaolinite was modelled as a single slab

[2]The lateral cell dimensions for the homogeneous simulations were obtained from a 0.01 K NPT simulation of a proton ordered configuration of hexagonal ice. In the case of the heterogeneous simulations, the lateral cell dimensions are constrained to be commensurate with the kaolinite slab. The kaolinite structure used was based upon the experimental structure of Bish [99], with the α and γ angles altered slightly to make the cell orthogonal.

and 192 water molecules were placed on the hydroxyl-terminated (001) face, creating a solid-liquid interface, whilst leaving a liquid-vapour interface. Due to the presence of the kaolinite substrate, the lateral dimensions were *ca.* $15.5 \times 17.9\,\text{Å}^2$, slightly larger than in the homogeneous case. To ensure that the kaolinite slab did not drift, one of the silicon atoms was fixed throughout the simulations.

To propagate the dynamics, the velocity Verlet algorithm was used with a timestep of 2 fs. Simulations were performed in the canonical ensemble and the temperature was controlled using a Nosé-Hoover chain of length 10 and a temperature coupling constant of 0.5 ps. Both systems were equilibrated at 300 K for 2 ns from which initial configurations for the production runs were sampled. For the production simulations, the systems were quenched to 220 K (i.e. approximately 30 K supercooled) and ran for the order of 1 μs or until nucleation was observed. The water geometry was maintained using the `SETTLES` algorithm [101], whereas the `P-LINCS` algorithm [102] was used to constrain the O–H bond in kaolinite. All simulations were performed using the `GROMACS 4.5` simulation package [65].

4.4 Results and Discussion

In total, 27 heterogeneous simulations (observing 10 nucleation events) and 30 homogeneous simulations (observing 9 nucleation events) were performed. Movies of some of these are available on the enclosed CD-ROM and online [103] (see also Appendix C). Before doing any detailed analysis one trend was immediately clear: on the kaolinite exclusive formation of hexagonal ice was observed, whereas in the homogeneous simulations, generally a mixture of hexagonal and cubic stacking patterns formed. In all but one of the homogeneous simulations, at least half of the ice formed consisted of cubic sequences with the bilayers parallel to the liquid/vapour boundary. Only one simulation resulted in solely hexagonal ice. This is qualitatively consistent with X-ray diffraction data and Monte Carlo simulations performed by Malkin *et al.*, which demonstrated that the homogeneous nucleating phase is stacking disordered (denoted as ice I_{sd}), consisting of roughly equal numbers of cubic and hexagonal sequences [104, 105]. This mixture of cubic and hexagonal layers is also consistent with previous simulation studies [69, 106]. Furthermore, when ice forms homogeneously a variety of crystal orientations within

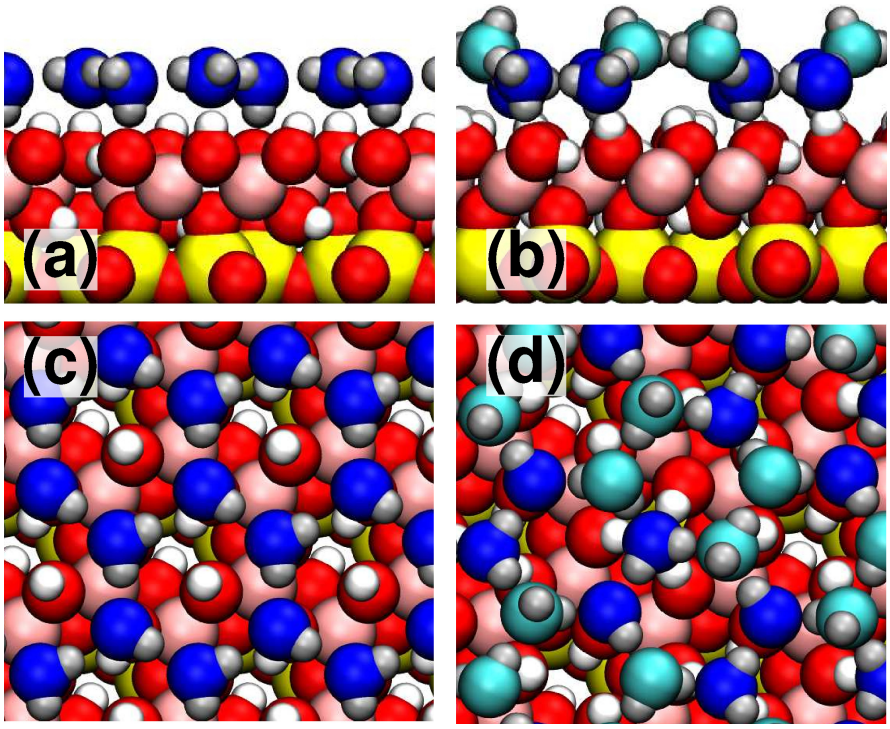

Figure 4.2: Diagram of ice-like structures at the kaolinite surface. In panel (a) a side view of the basal face of ice bound to kaolinite in the "H-down bilayer" configuration [66] is shown. All water molecules bind with similar heights from the surface. Panel (b) shows a side view of the prism face bound to kaolinite. In this structure, the water molecules come in high-lying (light blue) and low-lying (dark blue) pairs. Note that the prism face structure donates hydrogen bonds to the surface, as well as having 'dangling' hydrogen bonds pointing away from the surface (these dangling hydrogen bonds are absent in the basal face structure). Panels (c) and (d) show top views of the basal and prism face structures, respectively.

the simulation cell is seen, whereas when ice forms on the kaolinite, growth along the prism face of ice and not the basal face is always observed i.e. the ice bilayers grow perpendicular to the kaolinite slab. Figure 4.2 shows diagrams of the basal and prism faces at the kaolinite surface. The observation that the prism face nucleates at kaolinite is interesting, as it means that the pseudo-hexagonal arrangement of –OH groups at the kaolinite surface are not acting as a template for the basal face of ice.

From visual inspection of the ice-forming trajectories it was noticed that, during the nucleation event, considerable rearrangement of the water molecules always seemed to occur in the second water layer above the kaolinite surface (note that this statement does not preclude any rearrangement occurring in the first or third layers). To provide evidence for this observation, how the density of water

varies with height along the z-direction during the transition has been measured. In Figure 4.3 this analysis is presented for a single heterogeneous and homogeneous simulation along with the corresponding snapshots. First of all, it is seen that the supercooled liquid (shown at 55 ns) has an extremely sharp and intense density peak at the kaolinite surface, as well as a pronounced, but broader, second peak.[3] After 61.5 ns the nucleation event has occurred and it is seen that the intensity of the first peak has decreased slightly, although it still remains much higher than anywhere else in the system. It is also seen that the second peak has started to split (highlighted in yellow), indicative of an ice-like layer forming. It is only after this change in density in the second layer that one sees the first layer transform fully to ice. This is in contrast to the homogeneous case, where the density in the supercooled liquid is essentially uniform and the nucleation event seems to occur by two or three layers concurrently forming ice. In both the homogeneous and heterogeneous scenarios, once the initial nucleation event has occurred the growth of ice then proceeds, with a quasi liquid-like layer remaining at the water/vapour interface, consistent with previous simulation studies [69, 106–108]. It is noted that the observed changes away from the surface have striking similarity with the previously reported 'collective mechanism' for ice growth along non-basal faces at temperatures below 240 K [109, 110].

To investigate these structural changes away from the surface further, for each heterogeneous nucleation event observed the density difference:

$$\Delta\rho(z) = \rho(z) - \langle\rho_{\mathrm{liq}}(z)\rangle \tag{4.5}$$

has been computed, where $\rho(z)$ is the instantaneous water density at a height z and $\langle\rho_{\mathrm{liq}}(z)\rangle$ is the water density at a height z averaged over supercooled liquid configurations. The results are presented in Figure 4.4 (for reference, panel (b) corresponds to the heterogeneous nucleation event presented in Figure 4.3). It can clearly be seen that in all instances, just after the onset of nucleation, there is a change in the density of the second (and often the third) water layer that is

[3] This high density peak is also a feature of water at kaolinite at 300 K, which is shown in Appendix C, along with a density profile from a DFT-MD simulation at 330 K that also exhibits such a peak.

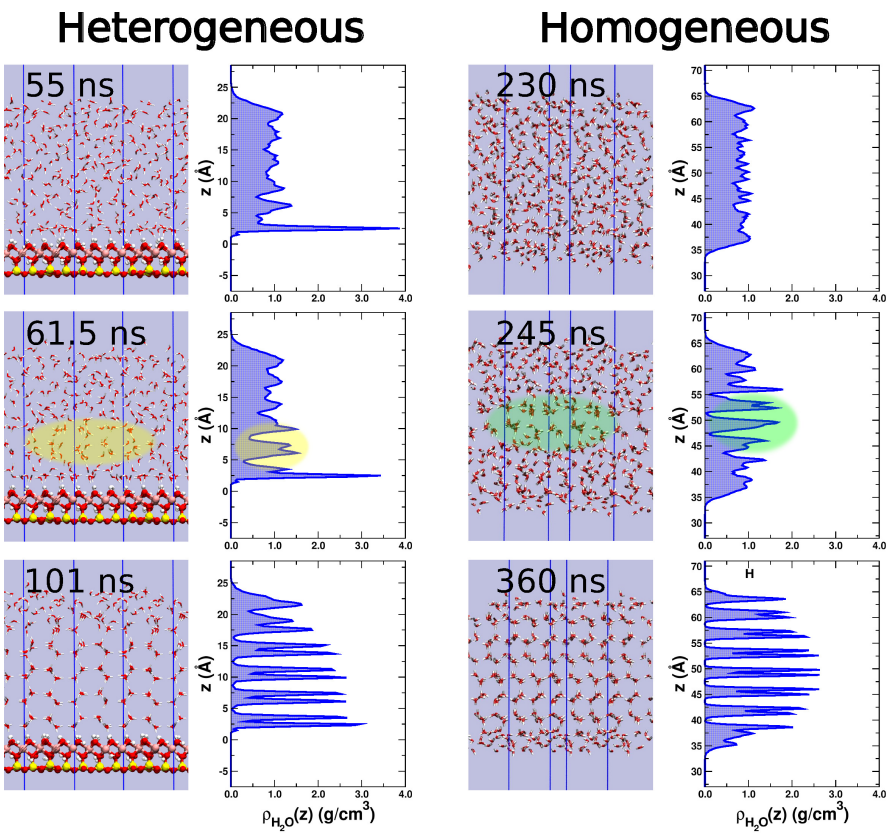

Figure 4.3: Snapshots and water density profiles for a homogeneous (right) and heterogeneous (left) nucleation event. In the presence of kaolinite, the supercooled water (55 ns) has a high density peak corresponding to the first contact layer. There is also a noticeable second peak, but this is far less intense and much broader. After 61.5 ns, nucleation has started. There is a slight reduction in density of the first density peak, but this is still much higher than anywhere else in the system. Rearrangement of water molecules in the second layer associated with a split in the density peak (highlighted in yellow), is also seen and is indicative of an ice-like layer forming. By 101 ns the first contact layer has fully transformed to ice and the density is similar to that observed in the rest of the system. Note that it is the prism face of ice exposed to the kaolinite surface and that only hexagonal ice is observed. In contrast, for the homogeneous slab one sees a fairly uniform density profile in the supercooled regime (230 ns). A mixture of hexagonal and cubic stacking is also seen. In this instance, the initial nucleation event (highlighted in green) leads to a cubic stacking arrangement. The densities are averages over a 2.5 ns interval centred at the specified time. The colour scheme is the same as Figure 4.1.

comparable to the changes observed in the first layer. In none of the simulations is it observed that the first layer fully transforms to ice without this signature splitting of the second layer density.

It is important to consider how significant the changes in density away from the surface are in the ice nucleation mechanism; after all, one may argue that these are just a consequence of the initial changes seen in the first layer and are merely

indicative of ice growth rather than playing a role in the nucleation mechanism itself. A committor analysis on the heterogeneous trajectory presented in Figure 4.3 (and panel (b) in Figure 4.4) has been performed, using the `CHILL` algorithm of Moore *et al.* [111] to monitor ice formation. This was done by choosing different configurations along this trajectory and starting 10 new trajectories with random velocities drawn from the Maxwell-Boltzmann distribution. Results from three starting configurations are presented in Figure 4.5, where one can clearly define a pre- and post-critical region (initial configurations from 62.5 ns and 70.0 ns of the initial trajectory, respectively). In between these two regimes, however, one does not see an expected 50:50 split of trajectories going on to reach the liquid and ice states, rather it is seen that some definitely go to ice, some definitely go to liquid but some trajectories stay somewhere in between, even over fairly long timescales (*ca.* 50 ns). As the cost of this committor analysis is high, no further attempt to refine the committor anaysis further has been made and one can remain satisfied that the configuration sampled at 65.0 ns is a reasonable representation of the 'transition region'. What is relevant to the discussion regarding the density changes in the second layer is that $\Delta\rho(z)$ shown by the red line shown in Figure 4.4(b) corresponds to the configuration sampled at 62.5 ns i.e. the splitting in the second peak for this trajectory occurs in the pre-critical regime, indicating that these structural changes are part of the nucleation mechanism rather than a feature of growth. A similar analysis for the trajectory in Figure 4.4(d) has also been performed, which is presented in Appendix C, along with movies showing how $\Delta\rho(z)$ varies during the nucleation event (provided on the enclosed CD-ROM and online [103]).

It is interesting to attempt to explain some of these observations. To help understand why the formation of ice with its prism rather than basal face exposed to the kaolinite surface is seen, the dependence of adsorption energy of ice on the number of ice-like layers, for both the prism and basal faces bound to the kaolinite has been investigated (details of these calculations are given in Appendix C). The results of this analysis are presented in Figure 4.6, where the data in terms of adsorption energy per water molecule and adsorption energy per conventional unit cell of kaolinite are shown. When only the first contact layer is present, the

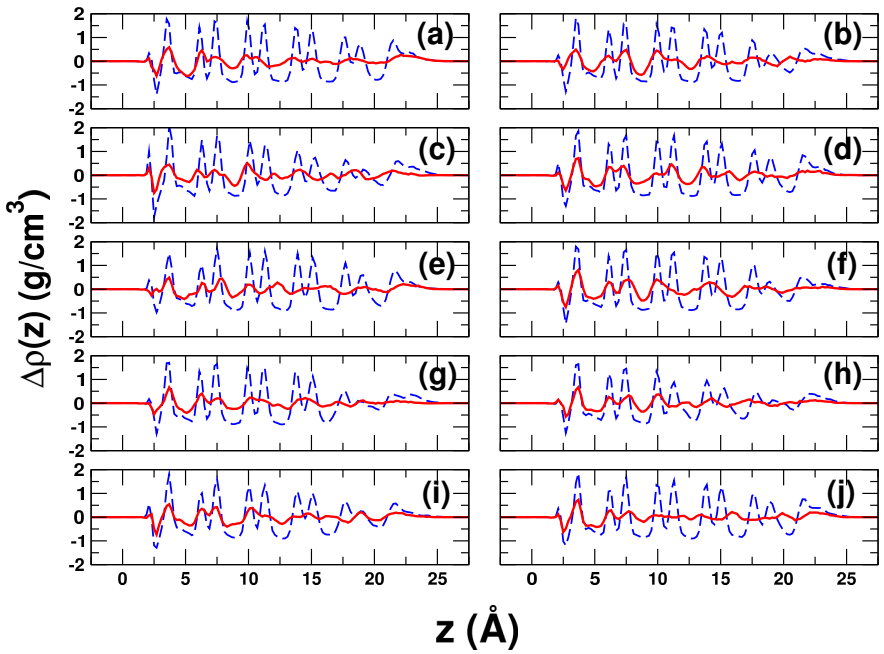

Figure 4.4: Water density difference profiles for all heterogeneous nucleation events. Each panel (a-j) shows an independent nucleation event. The quantity plotted is $\Delta\rho(z)$ as defined by Equation 4.5. The red solid line shows $\Delta\rho(z)$ at a time just after the onset of nucleation and the blue dashed line shows $\Delta\rho(z)$ at a later time after ice has grown. In all cases, it is seen that there are density changes in the second layer (just below 7.5 Å) of a similar size to those in the first layer, before ice goes on to form fully (noticeable changes in the third layer are also often observed). In the case of (b), from a committor analysis it is known that the red line corresponds to a pre-critical configuration. The displayed densities are averages over 2.5 ns.

basal face of ice is more strongly bound than the prism face by approximately 15 meV/H_2O. As soon as one goes beyond the first layer, however, the prism face becomes more stable, with the difference becoming more pronounced as more layers are added. The prism face also binds with a higher coverage than the basal face (5.33 vs. 4 H_2O per conventional unit cell) meaning that the prism face is more stable per unit cell of kaolinite, independent of the number of ice-like layers. To understand these differences, it is useful to examine the structure of the ice-like layers when binding through the prism and basal faces, which is shown in Figure 4.2. Here it can be seen that the water molecules in the basal face structure bind with similar heights from the kaolinite, with half the molecules donating one hydrogen bond to the surface and the other half accepting a hydrogen bond from the kaolinite whilst donating two hydrogen bonds to other water molecules

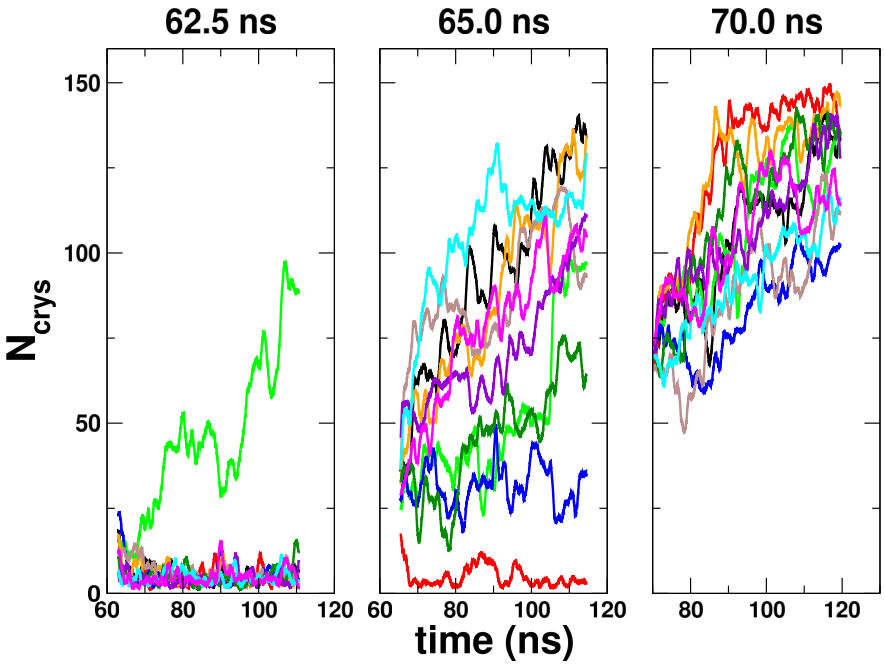

Figure 4.5: Committor analysis from one of the heterogeneous ice nucleation trajectories. Results here are shown for initial configurations sampled at 62.5 ns, 65.0 ns and 70.0 ns from the initial ice forming trajectory. 10 independent trajectories were started from each configuration by giving the particles random velocities sampled from a Maxwell-Boltzmann distribution. By monitoring the number of water molecules defined as being ice by the CHILL algorithm [111] N_{crys}, one is able to determine whether or not ice forms. It can clearly be seen that at 62.5 ns the system is in a pre-critical regime and by 70 ns all trajectories continue to form ice. At 65.0 ns one does not see all trajectories form ice or liquid, but that some stay somewhere in between the two states, even over the *ca.* 50 ns timescale. Results are presented as running averages over a 1 ns interval.

(the "H-down bilayer" structure as described in Reference [66]).[4] By adopting this structure, the water molecules maximise their bonding to the kaolinite and maintain good hydrogen bonding between each other, giving a large overall adsorption energy for the first layer. This structure, however, saturates all hydrogen bonds leaving no 'dangling' hydrogen bonds that water molecules in above layers can bind to and consequently, the adsorption energy rapidly becomes less exothermic as other layers are added. This finding is consistent with previous findings from a DFT study [66], as well as the experimental observation that the availability of dangling hydrogen bonds determines the multilayer wetting behaviour of water on metal and metal oxide surfaces [112]. On the other hand, the prism face binds with a somewhat more corrugated configuration, with the water molecules

[4]The "H-up bilayer" structure was also tested (see Reference [66]), but this was always less stable than the "H-down" structure and so has been omitted for clarity.

coming in high-lying and low-lying pairs. One of the molecules in the low-lying pair donates one hydrogen bond to the kaolinite whilst its partner accepts hydrogen bonds from the kaolinite. The high-lying pairs bridge the low-lying pairs through hydrogen bonds, with the important feature that one of these high-lying molecules has an OH bond directed away from the surface i.e. the prism face exhibits dangling hydrogen bonds. The fact that half of the molecules come in high-lying pairs means that the adsorption energy per water molecule of the first layer is less for the prism face than it is for the basal face, but the ability of the prism face to donate and accept hydrogen bonds to both the surface and the above water layers means that it becomes more stable as the number of water layers increases. The adsorption energies of the first and second layers with DFT using the Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional [113] have also been computed (full details of these calculations are given in Appendix C). Although agreement is not exact between our force field setup and PBE (which should not be taken as a benchmark) the trend that the prism face becomes more stable than the H-down bilayer upon adsorption of a second layer of ice is still seen. This suggests that this observation is not an artifact of the choice of force field.

In Figure 4.6 the average adsorption energy of the first layer from 25 configurations selected from the supercooled liquid is also shown.[5] On a per molecule basis, the liquid layer is less stable than either of the ice-like structures, but from the right hand panel of Figure 4.6 it is seen that per unit cell of kaolinite, the liquid layer is slightly more stable. This result may help explain the observed density changes away from the surface during the nucleation process. If one draws an analogy to the grand canonical ensemble, one may consider the first water layer as a subsystem that is able to exchange heat and particles with the bulk liquid above.[6] In the supercooled state, therefore, there will be some pseudo-equilibrium number of water molecules in the first layer, which has been measured to be 5.61 H_2O/unit cell (c.f. 5.33 H_2O/unit cell for the prism face). Thus, although the

[5]Water molecules are defined as being in the first layer if their oxygen atoms are within 5 Å of the average height of the kaolinite surface oxygen atoms. This corresponds to a cutoff that is approximately halfway between the first and second ice layers.

[6]It must be emphasised that this is an analogy to the grand canonical ensemble and that the first water layer and the above liquid are strongly coupled.

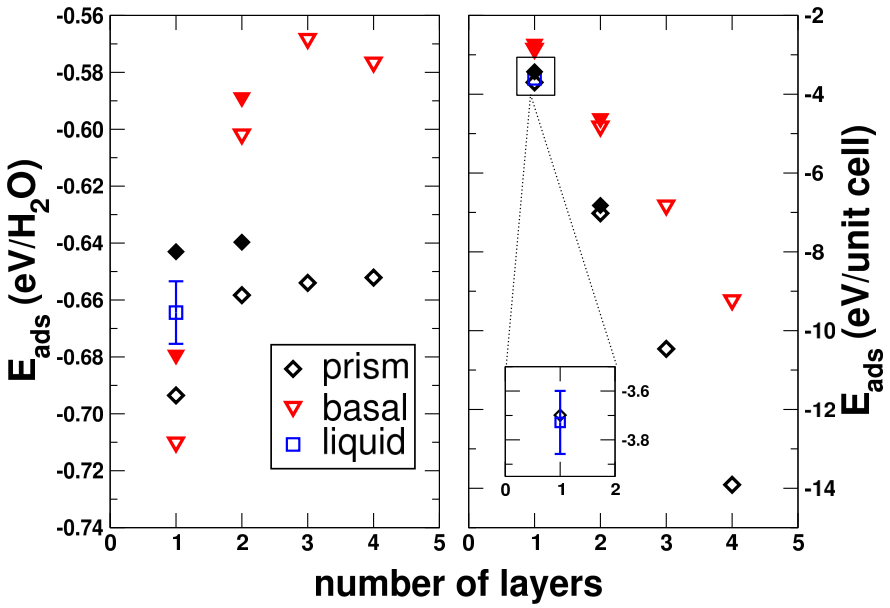

Figure 4.6: Variation of the adsorption energy (E_{ads}) of ice to kaolinite as the number of ice layers changes. The black diamonds show results for ice binding to kaolinite through its prism face, whilst the red triangles show results for ice binding through its basal face. Filled symbols show results from DFT calculations. The left panel shows the adsorption energy calculated per water molecule, whereas the right panel shows the adsorption energy per conventional unit cell of kaolinite. For the first contact layer on its own, the adsorption energy per water molecule is stronger for the basal face than the prism face, but upon adsorption of other layers, the prism face structure becomes significantly more stable. When ice binds through the prism face, the coverage of water molecules is higher than when it binds through the basal face, meaning that the adsorption energy per unit cell of kaolinite is more stable for the prism face independent of the number of adsorbed layers. Data for the first liquid layer is also shown (the bars indicate estimates of the thermal fluctuations). On a per molecule basis, this is less stable than the ice-like structures, but is more stable per unit cell of kaolinite.

adsorption energy per water molecule is stronger for the first layer of ice, on average more water molecules are present in the first liquid layer leading to an overall stabilisation. For ice to form and persist at the surface, it is therefore required that the average number of water molecules at the surface decreases. In keeping with the analogy to the grand canonical ensemble, this amounts to a need for a change in the chemical potential of the reservoir of water molecules above the first layer, which manifests itself as the structural changes away from the surface discussed previously. It can also be seen from the right panel of Figure 4.6 that the adsorption energy of the prism face per unit cell is within our estimate of the thermal fluctuations from the average liquid value. This may be one of the reasons for kaolinite's good ice nucleating ability.

Finally, it is important that the role of finite size effects is mentioned. It has been attempted to perform these simulations with system sizes doubled in the lateral dimensions (768 water molecules, cutoff for interactions extended to 9 Å) but no nucleation was observed in a total simulation length of 15.5 μs over a temperature range spanning of 190–220 K. Simulations were also performed for 2 μs at 240 K using the TIP4P/ice water model [114], which has a melting point similar to experiment, but still no nucleation was observed. This discrepancy can be explained by the fact that in the small systems, there is a self interaction of the growing ice nucleus with its periodic images that lowers the interfacial free energy cost of nucleation. Using the same cell as for the heterogeneous simulations (i.e. lateral dimensions of *ca.* 15.5×17.9 Å^2 without the kaolinite slab), 14 homogeneous simulations have also been performed. No nucleation events were observed. Although it would have been desirable to have observed nucleation in these simulations, so that we could have compared homogeneous and heterogeneous rates, one pleasing aspect of this last null result is that it means that the heterogeneous nucleation results presented earlier are not completely dominated by finite size effects. As it is possible to routinely observe nucleation in this cell when the kaolinite slab is present, but not homogeneously, one is left to conclude that kaolinite significantly enhances the rate. It is not currently possible, however, to go beyond this qualitative level.

As a final test of the finite size effects, the lateral cell dimensions were doubled and a configuration from one of the heterogeneous simulations, replicated in both dimensions to fill the larger cell, was used as an initial configuration. Taking the configuration that was determined to be representative of the transition region from the committor analysis as a 'seed' configuration for the larger cell, the growth of ice in the same manner as the small cells was still observed. The fact that growth is seen and not a collapse of the crystal suggests that the prism face is stable on the hydroxyl-terminated (001) kaolinite face and is not solely stabilised by periodic boundary effects present in the small cells.

4.5 Conclusions

Ice nucleation in thin water films, both homogeneously and heterogeneously in the presence of a kaolinite slab have been investigated, using regular molecular dynamics simulation. Many simulations on the order of one microsecond have been performed, with many nucleation events observed. In agreement with previous simulation studies and recent experiments, in the case of homogeneous nucleation, a mixture of cubic and hexagonal arrangements is seen. Contrary to expectation, at the kaolinite surface growth along the prism face of ice is always seen, suggesting that the source of kaolinite's good ice nucleating ability does not lie with its good epitaxial match with the basal face of ice. By monitoring the density of water above the kaolinite slab during the nucleation event, it is seen that changes in the second water layer appear crucial to the nucleation mechanism. The growth of the prism face rather than the basal face is due to the ability of the former to bind favourably to both the surface and water layers above, as well as having a higher coverage. The observed structural changes away from the surface have been explained as allowing the average number of water molecules in the first layer to decrease, which subsequently allows the remaining water molecules to form the favoured ice-like structure. It has, however, been noted that finite size effects are non-negligible in these simulations, with no nucleation observed upon moving to bigger cells. Nevertheless, the fact that homogeneous nucleation is not observed in the cell size used for the heterogeneous nucleation simulations suggests that the results on kaolinite are not entirely dominated by the finite size effects. This result also shows that kaolinite is a potent ice nucleating agent.

The fact that the pristine kaolinite surface promotes the growth of the prism face over the basal face may have consequences for the macroscopic crystal structure of ice that forms. Ice exhibits a complex habit diagram [115] and as the surface cleavage energies of the prism and basal faces are very similar [116] it is possible that different heterogeneous ice nucleating agents could tip the balance to favour different ice habits under the same conditions. As the macroscopic structure of an ice crystal can affect its light scattering properties, understanding the effect of ice nucleating agents may be important for global climate models.

Given the finite size effects, it would be highly desirable to implement a free energy method that could definitively probe the heterogeneous nucleation mechanism proposed here. Even with the current state-of-the-art in free energy methods and advanced sampling techniques, freezing water is still likely to be difficult. The reason for this is that slow dynamics offered by the hydrogen bonding network present in supercooled water makes it very difficult for methods such as umbrella sampling and metadynamics to equilibrate the system as it is pushed along the chosen order parameter [117]. Furthermore, advanced sampling techniques that exploit natural dynamics, such as transition path sampling [118] or forward flux sampling [119, 120] are likely to suffer as the actual transition time is relatively long (tens of nanoseconds, as seen in Figure 4.5), which may make sampling computationally prohibitive. One way to circumvent this problem is through the use of a coarse grained potential such as the mW model [80] that, by treating hydrogen bonding in a mean-field sense, reduces the complexity of the underlying potential energy surface and results in faster dynamics. This has already been used to good effect with both direct molecular dynamics (see e.g. Reference [121]) and forward flux sampling [122] for homogeneous nucleation. This approach is unlikely to work in the case of heterogeneous nucleation on substrates such as clays, however, where electrostatics are dominant. How to proceed in such cases is at present unclear and, given the industrial and environmental implications of ice formation, is an issue that urgently needs to be addressed. In the remaining two chapters, however, the coarse grained approach will be embraced, as this is currently the most efficient way to gain insight into heterogeneous nucleation. We will also see an implementation of the transition path sampling technique, which might one day be feasible to use with all-atom potentials.

Chapter 5

Nanoscale Control of Ice Formation

5.1 Coarse Grained Water

This chapter will also use MD (see Section 4.1) to investigate heterogeneous ice nucleation, but rather than the all-atom potentials like Chapter 4, a coarse grained approach will be used. The details of this method will quickly be summarised.

The coarse grained approach adopted here is the mW model of water [80]. The key observation that led to the development of this model is that water is a tetrahedral liquid whose tetrahedrality is intermediate between liquid carbon and silicon. It therefore seems reasonable to try and model water as a single tetrahedral "atom" rather than trying to mimic the charge distribution around a water molecule. In this spirit, Moore and Molinero [80] took the Stillinger-Weber potential for silicon [123] and reparametrised it to reproduce the properties of bulk liquid water. The Stillinger-Weber potential is given by:

$$U_{\mathrm{SW}}(\mathbf{r}^N) = \sum_i \sum_{j>i} \phi_2(r_{ij}) + \sum_i \sum_{j\neq i} \sum_{k>j} \phi_3(r_{ij}, r_{ik}, \theta_{ijk}), \tag{5.1}$$

where:

$$\phi_2(r) = A\epsilon_{\mathrm{SW}} \left[B \left(\frac{\sigma_{\mathrm{SW}}}{r} \right)^4 - 1 \right] \exp\left(\frac{\sigma_{\mathrm{SW}}}{r - a\sigma_{\mathrm{SW}}} \right); \tag{5.2}$$

and

$$\phi_3(r, s, \theta) = \lambda\epsilon_{\mathrm{SW}} \left[\cos\theta - \cos\theta_0 \right] \exp\left(\frac{\gamma\sigma_{\mathrm{SW}}}{r - a\sigma_{\mathrm{SW}}} \right) \exp\left(\frac{\gamma\sigma_{\mathrm{SW}}}{s - a\sigma_{\mathrm{SW}}} \right). \tag{5.3}$$

The parameters $A = 7.049556277$, $B = 0.6022245584$, $p = 4$ and $\gamma = 1.2$ give the form and scale to the potential and the reduced cutoff $a = 1.8$ ensures that all terms in the potential go continuously to zero at a distance $a\sigma_{\mathrm{SW}}$. The three-body term ϕ_3 penalises configurations of triplets of water molecules that deviate from the tetrahedral angle $\theta_0 = 109.47°$. The remaining parameters are λ, ϵ_{SW} and σ_{SW}, which control the tetrehedrality, energy- and length-scale, respectively. By parameterising $\lambda = 23.15$, $\epsilon_{\mathrm{SW}} = 6.189\,\mathrm{kJ/mol}$ and $\sigma_{\mathrm{SW}} = 2.3925\,\text{Å}$, a model that captures the structure and energetics of bulk liquid water, ice and amorphous solid water reasonably well can be obtained. It also captures a number of the anomalous properties of liquid water.

Reducing a water molecule into a single interaction site might seem like a drastic measure. So what is to be gained from doing this? First, the computational cost is lower as the number of interaction sites has been reduced and, without explicit hydrogen, longer time steps can be used for MD simulation.[1] These gains, however, are not enough to sacrifice the detail contained in all-atom potentials. The principle reason for using mW is because the potential energy surface is much smoother than that obtained with all-atom potentials, meaning that configuration space is sampled much more easily.[2] Consequently, mW undergoes nucleation more readily than, say, the TIP4P/2005 water model used in the previous chapter. This gain in statistical sampling of nucleation for larger simulation sizes warrants such a drastic simplification.

5.2 Introduction

In the previous chapter, we saw the first simulations of heterogeneous ice nucleation where both the atomic structure of the water and the INA were taken into account. Although this provided fresh insight into heterogeneous ice nucleation, it was noted that finite size effects were affecting the simulations. In order to remediate this, the coarse grained mW potential [80] will be used in this chap-

[1] Reports of speed-ups of a factor 180 compared to all-atom potentials are, however, fanciful. With optimised Fast Fourier Transform and parallelisation algorithms, pairwise additive all-atom potentials with electrostatics have been found to be approximately 10 times slower than the three-body mW potential.

[2] This is also reflected by a diffusion coefficient that is approximately 3 times too large compared to experiment.

ter, which allows for simulations of ice nucleation with a far greater number of molecules than the all-atom potentials used previously. Of course, this comes at the cost of a loss of detail (e.g. the significance of the dangling hydrogen bond seen in Chapter 4 cannot be captured). Nevertheless, much can be gained from the coarse grained approach if one is willing to consider general trends rather than specific systems. This chapter therefore sees a return to the simplified surface potentials analogous to Chapter 3.

As noted in Chapter 1, in contrast to fields such as chemical catalysis [22] and materials design [38], there is currently no comprehensive set of design principles in terms of molecular 'descriptors' for making new substances to control ice formation. Put more simply, we do not know which are the relevant microscopic properties of a material that determine its macroscopic ice nucleating efficiency. We also saw in Chapter 2 that in the ice formation community, the so-called 'requirements' for a good ice nucleating agent (INA) have often been discussed, such as the requirement for a good crystallographic match to ice and the ability of water to chemically bond to the surface of the particle (i.e. hydrophilicity) [1]. Such criteria have neither served as a full set of guidelines to identify good INAs [7], nor have they aided the systematic improvement of ice nucleation inhibitors or promoters (for example, experimentally there has been disagreement regarding the role of hydrophilicity [124–126]). As a community, we are either faced with the prospect of relying on experiments to determine the efficacy of INAs on a case-by-case basis, or we can try and rationalise their behaviour by elucidating the underlying molecular processes that control heterogeneous ice nucleation.

In this chapter, results from MD simulations are presented where heterogeneous ice nucleation is probed directly in the presence of three types of model nanoparticle (NP) of varying topography (each approximately 2.5 nm in diameter and completely immersed in water): a smooth, structureless sphere, hereafter referred to as "smooth"; a face centred cubic (FCC) hemisphere that exposes an hexagonal (111) face as its principal facet, referred to as "FCC-111"; and a graphene nano flake, referred to as "GNF". As well as allowing changes in curvature to be probed, the choice of these three NPs also enables evaluation of the effects of surface structure. In particular, the GNFs present surfaces that are geo-

metrically flat and that, due to the short carbon-carbon bond lengths, are unable to act as an hexagonal template for the water molecules (see e.g. Chapter 3 and Reference [19]). On the other hand, the (111) surface of the FCC-111 NP, while also geometrically flat, presents distinct adsorption sites for the water molecules in the contact layer, which can provide a template for the basal face of ice. The hydrophilicity of each of these NPs is varied by altering the interaction strength with water and by performing a large number of simulations for each system, quantitative comparison of nucleation rates is possible. Not only is it possible to control the rate of ice formation, spanning inhibition to promotion relative to homogeneous nucleation, but it is also shown that the dependence of the rate on surface hydrophilicity differs between these NPs. It will also be demonstrated how one can use the understanding of the mechanism observed on the FCC-111 NP to effect enhanced ice nucleation. Ways in which the findings of this simulation study can be tested experimentally will be discussed and it is suggested that this study paves the way for the control of ice formation driven by fundamental insight.

5.3 Simulation Methods

All simulations were performed using the `LAMMPS` simulation package [127] and the coarse grained mW model for water [80]. The velocity Verlet algorithm was used to propagate the equations of motion of the water molecules, using a 10 fs time step. 2944 mW molecules were used in all simulations. Previous simulation studies have suggested that the critical ice nucleus varies from ∼10 water molecules at 180 K [72] to ∼85–265 at 220 K [74, 122] giving one confidence that the simulations should not be subject to serious finite size effects. Periodic boundary conditions were applied in all three dimensions. Temperature and pressure were maintained using the Nosé-Hoover thermostat and barostat (with a chain length of 10) with relaxation times of 1 ps and 2 ps respectively. For each system, a 100 ns trajectory was first performed at 290 K and 1 bar, from which initial configurations were drawn (different initial configurations were separated by at least 5 ns in the high temperature trajectory). At the start of the nucleation simulation, velocities for the water molecules were drawn randomly from a Maxwell-Boltzmann distribution to give an initial temperature of 205 K. Simulations were stopped after 500 ns if

nucleation did not occur. To detect 'ice-like' molecules, the `CHILL` algorithm of Moore *et al.* [111] has been used.

To model the smooth NP, the 'expanded Lennard-Jones' potential available in `LAMMPS` was used:

$$U_{\mathrm{LJ}}(r) = 4\epsilon_{\mathrm{LJ}}\left[\left(\frac{\sigma_{\mathrm{LJ}}}{r-\Delta}\right)^{12} - \left(\frac{\sigma_{\mathrm{LJ}}}{r-\Delta}\right)^{6}\right], \tag{5.4}$$

where ϵ_{LJ} and $\sigma_{\mathrm{LJ}} = 0.234\,\mathrm{nm}$ are the usual Lennard-Jones parameters that control the well-depth and range of interaction, respectively, r is the distance between a water molecule and the centre of the nanoparticle, and $\Delta = 1\,\mathrm{nm}$ is a parameter that controls the size of the nanoparticle (the effective hard-core radius of the nanoparticle is $\Delta + \sigma_{\mathrm{LJ}}$). Interactions were truncated at a distance $\Delta + 0.753\,\mathrm{nm} \approx \Delta + 3.2\sigma_{\mathrm{LJ}}$ from the centre of the NP. The value of ϵ_{LJ} defines the adsorption energy E_{ads} of a single water molecule to the smooth NP.

The FCC-111 NP was constructed of regular Lennard-Jones spheres (i.e. Equation 5.4 with $\Delta = 0$; $\sigma_{\mathrm{LJ}} = 0.234\,\mathrm{nm}$) placed on an FCC lattice with a lattice constant of 0.392 nm. The use of these parameters yielded contact layers at a height between 0.2–0.25 nm above the FCC-111 NPs which is in reasonable agreement with values obtained from density functional theory calculations of water clusters at metal surfaces [60]. The NP was constructed using the Atomic Simulation Environment package [128]: 6 atomic layers were used in the {1,0,0} family of directions; 9 layers in the {1,1,0} family; and 5 layers in the {1,1,1} family, except along the $(\bar{1},\bar{1},\bar{1})$ direction where no layers were used. This resulted in a 380 atom hemispherical FCC nanoparticle that exposes the (111) face as its principal facet. Interactions were truncated after 0.753 nm. The total energy after geometry optimisation of a single water molecule at the centre of the (111) face was used to define E_{ads}.[3] As the equations of motion for the atoms in the FCC-111 NP were not integrated (i.e. they were fixed) no interaction potential was defined between them. The small adsorbates used to modify the FCC-111 NP were also modelled using Lennard-Jones spheres, with $\epsilon_{\mathrm{LJ}} = 0.18\,\mathrm{kcal/mol}$

[3] As the internal energies of an isolated water molecule and NP are both zero, E_{ads} is simply the total energy of a single water molecule bound to the NP.

and $\sigma_{\mathrm{LJ}} = 0.18\,\mathrm{nm}$ (like the rest of the FCC-111 NP, the equations of motion were not integrated for the adsorbates either).

Unlike the FCC-111 NP, the interaction of water with the GNF was modelled using the two-body part of the Stillinger-Weber potential (see Equation 5.2) in the same manner as Lupi *et al.* [129, 130]. The GNF consisted of 217 carbon atoms, with a carbon-carbon distance of 0.142 nm. As in References [129, 130], $\sigma_{\mathrm{SW}} = 0.32\,\mathrm{nm}$ was used to define the range of the water-carbon interaction. The interaction strength was tuned by varying ϵ_{SW}. Again, the total energy after geometry optimisation of a single water molecule at the centre of the GNF was used to define E_{ads}, and no interaction was defined between the carbon atoms as their equations of motion were not integrated.

16 MD simulations, each with a different initial configuration, were performed at 205 K and 1 bar. Under these conditions, bulk liquid mW water is still metastable (as opposed to unstable) [121] but undergoes homogeneous nucleation on a timescale accessible to computer simulation such that statistically meaningful rates can be obtained. In order to gauge the effectiveness of these NPs as INAs, bulk homogeneous nucleation was also investigated. Although relatively simple, such a simulation protocol allows the systematic comparison of ice nucleation for these different systems. Examples of ice nucleation in the presence of these NPs are shown in Figure 5.1 and time-resolved snapshots are provided in Appendix E, along with movies on the enclosed CD-ROM. By monitoring the potential energy, one is able to determine the induction time to nucleation for each simulation and consequently, the probability $P_{\mathrm{liq}}(t)$ that a given system remains liquid after a time t from the start of the simulation. One is able to determine the ice nucleation rate R by fitting $P_{\mathrm{liq}}(t) = \exp\left[-(Rt)^{\gamma}\right]$, where $\gamma > 0$ is also a fitting parameter. A detailed description of the fitting procedure is provided in Appendix E.

5.4 Results and Discussion

Figure 5.1 shows the dependence of R on E_{ads} for the three types of NPs. Specifically, $\log_{10}(R/R_{\mathrm{hom}})$ vs $E_{\mathrm{ads}}/\Delta H_{\mathrm{vap}}$ has been plotted, where R_{hom} is the bulk homogeneous rate and $\Delta H_{\mathrm{vap}} = 10.65\,\mathrm{kcal/mol}$ is the enthalpy of vapourisation

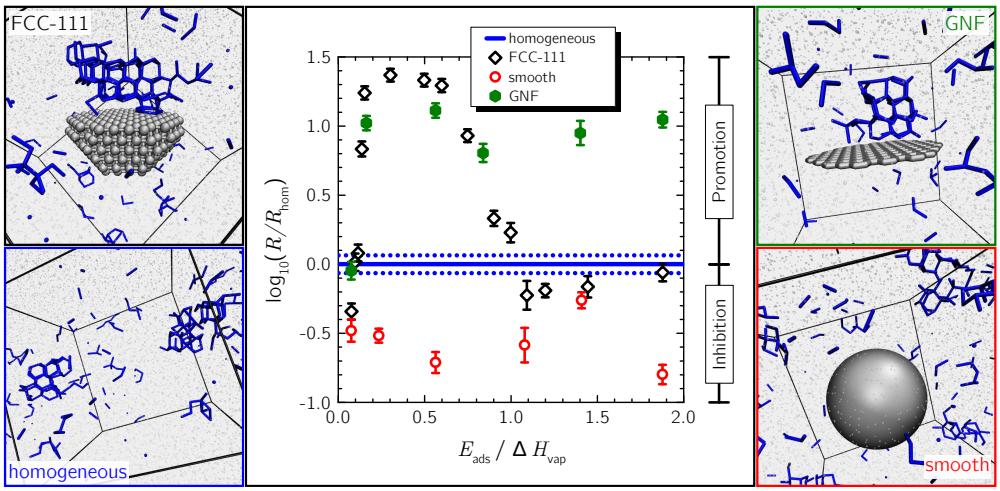

Figure 5.1: Controlling ice nucleation through surface topography and hydrophilicity. The central panel shows the dependence of the nucleation rate R on water monomer adsorption energy E_{ads} in the presence of the three NPs studied, plotted as $\log_{10}(R/R_{\text{hom}})$ vs $E_{\text{ads}}/\Delta H_{\text{vap}}$. As E_{ads} increases so too does the hydrophilicity. The solid blue line indicates the homogeneous nucleation rate (dotted blue lines indicate an uncertainty estimate). At very low values of E_{ads}, all of the NPs either inhibit or give rates similar to R_{hom}; this inhibition persists over the full range of E_{ads} for the smooth NP. The FCC-111 NP exhibits a maximum in R at $E_{\text{ads}}/\Delta H_{\text{vap}} \approx 0.4$ before falling to $R \leq R_{\text{hom}}$ at $E_{\text{ads}}/\Delta H_{\text{vap}} \approx 1.0$. With the exception of the weakest value of E_{ads}, the GNF promotes ice nucleation with a rate that remains relatively constant over the full range of E_{ads} (approximately 10 times faster than homogeneous nucleation). The snapshots show example nucleation events for the different systems: the NPs are shown in silver and molecules identified as ice-like are shown in blue and the water molecules that remain are indicated by small grey dots. Time-resolved snapshots of these nucleation events are provided in Appendix E.

of bulk mW water at 298 K [80]. A remarkably rich variety in the ice nucleating behaviour of the different NPs is observed, with all three NPs exhibiting a distinct dependence on E_{ads}. Particles are seen to both promote and, surprisingly, inhibit ice nucleation. Perhaps the most striking feature of Figure 5.1 is the trend exhibited by the data obtained with the FCC-111 NP. At low values of E_{ads}, the heterogeneous nucleation rate is approximately two times lower than R_{hom}, before rapidly increasing to reach a maximum at $E_{\text{ads}}/\Delta H_{\text{vap}} \approx 0.4$ that is nearly 25 times faster than bulk homogeneous nucleation. As E_{ads} is increased further, the rate steadily decreases until $E_{\text{ads}}/\Delta H_{\text{vap}} \approx 1.0$, beyond which the rate remains roughly constant and slightly below R_{hom}. In the presence of the GNF, with the exception of the lowest value of E_{ads} investigated, the rate is relatively invariant and is approximately ten times larger than R_{hom}. In the case of the smooth NP, inhibition is observed over the full range of E_{ads} investigated, with a rate that is

two to six times slower than homogeneous nucleation.

One can understand the rich behaviour exhibited in Figure 5.1 by examining the structure of the water molecules at the surface of the NPs during nucleation. Let us start with the FCC-111 NP. As the (111) surface of an FCC crystal exhibits hexagonal symmetry, one possible mechanism for heterogeneous ice nucleation is an in-plane template effect whereby the molecules in the contact layer form an hexagonal structure commensurate with the surface. Figure 5.2 (a) shows a typical nucleation event at the FCC-111 NP for $E_{\mathrm{ads}}/\Delta H_{\mathrm{vap}} \approx 0.3$, and indeed, it is clear that the water molecules at the surface do form such an overlayer that is acting as a template for an ice nucleus in the layers above. In cases where E_{ads} exceeds ΔH_{vap}, shown for example in Figure 5.2 (b) for $E_{\mathrm{ads}}/\Delta H_{\mathrm{vap}} \approx 1.2$, it is clear that the coverage of water molecules at the (111) surface has increased and the in-plane template effect that enhances ice nucleation is lost (nucleation is also seen to occur away from the surface in an homogeneous fashion). For the weakest E_{ads} investigated, $R \leq R_{\mathrm{hom}}$ is also caused by a loss of the in-plane template effect (the NP fails to significantly structure the water molecules).

Turning now to the GNF, for $0.1 < E_{\mathrm{ads}}/\Delta H_{\mathrm{vap}} < 1.4$ one also sees hexagonal overlayers forming at the surface, although this is no longer due to an in-plane template effect. An example is shown in Figure 5.2 (c) for $E_{\mathrm{ads}}/\Delta H_{\mathrm{vap}} \approx 0.16$, where it can be seen that the water molecules in the contact layer do not bind at specific sites at the surface. This lack of in-plane template effect becomes more obvious as E_{ads} is increased. For $E_{\mathrm{ads}}/\Delta H_{\mathrm{vap}} > 1.4$ the formation of hexagons in the contact layer is no longer observed, and a structure consisting predominantly of pentagons is seen (Figure 5.2 (d)). Despite this, nucleation still occurs through the formation of an hexagonal layer – now in the second layer of water molecules from the GNF – and the rate is unaffected. Recent simulation studies [129, 130] have attributed the ice nucleating ability of similar GNFs to the layering of interfacial water perpendicular to the surface. Initial analysis suggests that such layering is only part of the mechanism, the details of which are beyond the scope of this chapter and are discussed more thoroughly in Appendix D. Finally, in the presence of the smooth NP the water molecules in the contact layer do not form an ice-like structure and ice nucleation occurs away from the NP in an homogeneous manner

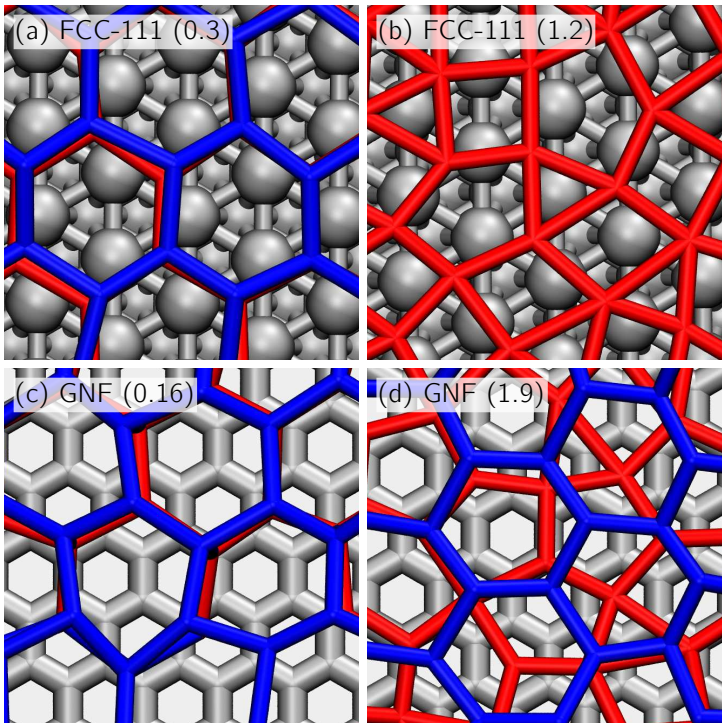

Figure 5.2: Snapshots of heterogeneous ice nucleation (only a portion of the surface is shown). Panels (a) and (b) are in the presence of the FCC-111 NP for $E_{\text{ads}}/\Delta H_{\text{vap}} \approx 0.3$ and 1.2, respectively. Water molecules in the contact layer are shown in red, otherwise the colour scheme is the same as Figure 5.1. The NP shown in (a) enhances ice nucleation by acting as a template for ice (the molecules are bound at specific adsorption sites and the ice is in registry with the substrate). As E_{ads} is increased, the coverage of molecules at the (111) surface increases and the in-plane template effect is lost (nucleation occurs away from the surface, so no ice-like molecules are seen in (b)). In (c) and (d) nucleation at the GNF is shown for $E_{\text{ads}}/\Delta H_{\text{vap}} \approx 0.16$ and 1.9, respectively. In (c), an hexagonal ice-like structure forms at the GNF, but the water molecules are not adsorbed at specific sites (no in-plane template effect). Although the water molecules in the contact layer do not form an ice-like structure in (d), nucleation still occurs through the formation of hexagons (in the second layer from the GNF).

for all E_{ads} (see Figure 5.1). It has been previously reported that a GNF with a radius of curvature of 2 nm hindered the growth of ice homogeneously [129]. Not only do the results for the smooth NP support this, but the broad range of E_{ads} over which this inhibition is observed, as well as the fact that the smooth NP is completely structureless, suggest that inhibition by surfaces of high curvature may be a general result (when the in-plane template effect is lost on the FCC-111 NP, which also exhibits regions of a similar curvature, inhibition has also been observed). It is interesting to note that several antifreeze proteins possess domains with curvatures similar to those of the smooth NP [131–136].

As has already been noted, the FCC-111 NP ceases to promote ice nucleation

for $E_{\text{ads}}/\Delta H_{\text{vap}} > 1$ due to the loss of the in-plane template effect, but the details of this observation warrant further discussion. When an ice-like hexagonal overlayer forms on the FCC-111 NP, such as in Figure 5.1 (a), it does so with sub-monolayer coverage at the (111) surface i.e. not all of the available adsorption sites are occupied by water molecules. These unoccupied adsorption sites on the (111) terrace are referred to as "excess" sites. When $E_{\text{ads}} > \Delta H_{\text{vap}}$ it becomes favourable for a water molecule to occupy a site on the surface, including the excess sites, rather than a position in the bulk liquid. By this rationale, if the density of available adsorption sites was lower, then the in-plane template effect (and the enhanced nucleation rate) may be preserved at higher values of E_{ads}. To this end, the FCC-111 nanoparticle has been modified by adsorbing small molecules, at the excess sites, which only have a weak interaction with water and the nucleation rate has been recomputed with $E_{\text{ads}}/\Delta H_{\text{vap}} \approx 1.2$. As seen in Figure 5.3, and consistent with the above discussion, this modified surface enhances ice nucleation by a factor of 50 compared to the unmodified surface. This is a clear demonstration of how the molecular insight into heterogeneous ice nucleation can be used to rationally design surfaces of different ice nucleating ability. Experimentally, this could be realised through adsorption of small molecules to the surface (e.g. carbon monoxide) or through surface alloying. In fact, the latter has already been observed to promote the formation of an hexagonal bilayer rather than an overlayer consisting of pentagons and heptagons seen at platinum (111) under ultra-high vacuum conditions [46, 54].

Although this chapter has focused on using computer simulation to study the fundamental aspects of heterogeneous ice nucleation, some suggestions for experiments can be made. For surfaces with hexagonal symmetry but a density of adsorption sites that is higher than required for ice, we have already suggested that ice nucleation could be enhanced through either molecular adsorption to the surface, or through surface alloying [54]. To test the inhibiting effect of high curvature, nanoparticles of different sizes but of the same composition could be added to water droplets – fullerenes provide an appealing possibility. The sensitivity of the nucleation rate on surface hydrophilicity could be tested using metal nanoparticles of gold or silica functionalised with organic molecules of varying hydropho-

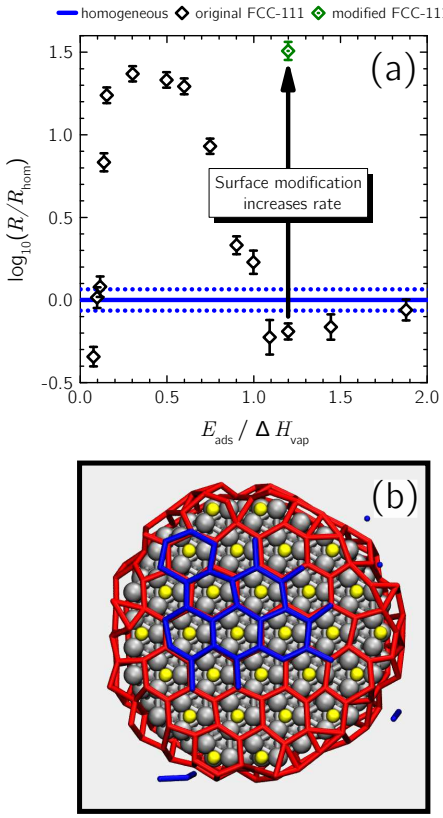

Figure 5.3: Panel (a) shows the enhancement of rate obtained after modifying the FCC-111 nanoparticle by adsorbing small molecules at the excess adsorption sites. Panel (b) shows a snapshot just after ice nucleation on the modified particle. The extra adsorbates used to modify the particle are shown by yellow spheres; otherwise the colour scheme is the same as Figures 5.1 and 5.2. By blocking the excess sites with these adsorbates, the in-plane template effect is recovered (c.f. Figure 5.2 (b)) and the nucleation rate is enhanced.

bicity. In addition to using well established methods such as the droplet freezing techniques familiar to the environmental sciences (see e.g. Reference. [17]), or neutron scattering, it may also be possible to exploit recent advances in femtosecond X-ray scattering techniques that have allowed real-time monitoring of homogeneous ice nucleation in micron sized water droplets [16]. Not only could such an experimental protocol be used to compare rates of ice nucleation in the presence of immersed NPs, but information regarding the impact of such NPs on the microscopic structure of the liquid should also be available.

5.5 Conclusions

In summary, computer simulations have been used to systematically compare heterogeneous ice nucleation rates in the presence of nanoparticles of varying

topography and hydrophilicity. It has been seen that distinct mechanisms of heterogeneous ice nucleation exist and that these mechanisms respond differently to changes in the hydrophilicity of the particles. Specifically, it has been seen that on planar surfaces that promote ice nucleation through layering of the interfacial water, the rate of ice nucleation is relatively insensitive to the hydrophilicity of the surface, whereas on surfaces that promote ice nucleation by acting as a template, the ice nucleating efficiency is lost if adsorption is too strong, due to a high coverage of water molecules destroying the in-plane template effect. Modification of the surface such that the coverage of water molecules is reduced recovers the in-plane template effect and enhanced nucleation can be achieved for strongly adsorbing surfaces.

Aside from clearly demonstrating how molecular level understanding of heterogeneous ice nucleation can be used to manipulate the rate of ice formation, this work also challenges the notion of 'requirements' for a good ice nucleating agent. For example, hydrophilicity is thought to be an important requirement of an INA, but it has been seen that different particles of the same hydrophilicity can either inhibit or promote ice formation. So rather than discussing 'requirements', it is suggested that it is more useful to identify 'descriptors' that actually correlate with the rate of ice nucleation. Although a comprehensive set of rules still requires further experimental and theoretical investigation, the results presented here suggest: (1) flatter surfaces are better at nucleating ice than curved surfaces; and (2) if the surface acts as a template, then one must tune either the density of adsorption sites, or the propensity of water to adsorb to the surface. We have seen in Chapter 1 that the use of descriptors has been successfully used in other fields, such as chemical catalysis [22]. Designing new catalysts for reactions such as methanation ($CO + 3\,H_2 \longrightarrow CH_4 + H_2O$) has relied upon the establishment of a Sabatier principle based on a computationally tractable quantity (in this case the dissociation energy of CO at the surface) [23]; whether such a principle exists in general, such that the ice nucleating ability of all materials can be described by just a few molecular decriptors, remains to be seen.

Chapter 6

The Role of Dynamic Heterogeneity

6.1 Transition Path Sampling

In the previous two chapters, ice nucleation has been investigated directly using molecular dynamics, first with an all-atom description and then using a coarse grained potential. The benefit of this approach is that there is minimal human influence on the results: once the interaction potentials are defined, the simulations are simply allowed to run and conclusions are drawn from whatever observations are made (i.e. these are 'simulation experiments'). There are drawbacks, however. Firstly, neglecting pathological issues such as finite size effects, the brute force approach is highly inefficient. Even with the coarse grained potential, much time is spent sampling the supercooled liquid state with only a fraction of the simulation contributing to the actual transition. Secondly, in order for the brute force approach to work, highly forcing conditions are required (i.e. very low temperatures) in order to simulate ice nucleation. As an example, taking the rate for bulk homogeneous nucleation for mW water at 220 K calculated by forward flux sampling [122], for the 2944 molecule system studied in Chapter 5 the characteristic nucleation time is approximately 5×10^{49} s (for comparison, the age of the universe is approximately 4×10^{17} s)! In other words, nucleation is a *rare event process*.

Rare event processes are common in chemistry and molecular physics – examples other than crystal nucleation include protein folding [137], autoionisation

of water [138] and pore translocation [139]. Consequently, many techniques have been developed to tackle the problem of sampling rare event processes. These methods can be loosely placed in one of two categories: (i) *free energy methods* and (ii) *path sampling techniques*. Two of the most common free energy methods are *metadynamics* [81, 82] and *umbrella sampling* [83]. With these methods, one defines a set of order parameters that distinguish the reactant and product basins (see Section 2.1.5), and biases the system along these order parameters. After correcting for the bias, one can then use Equation 2.26 to calculate the free energy profile along this chosen set of order parameters. The advantages of free energy methods are that they give direct information about the free energy barriers involved in the process of interest. The disadvantages are that they require an *a priori* knowledge of the reaction mechanism such that useful order parameters can be defined which for crystal nucleation, essentially amounts to guesswork. Path sampling methods, on the other hand, sample *trajectory space* rather than configuration space. The advantages of path sampling techniques is that they tend to be less sensitive to the choice of order parameter used and, as the Hamiltonian of the system is not modified, the dynamics remain unperturbed. Examples of such techniques include *forward flux sampling* [119, 120] and *transition path sampling* [118, 140] (multiple variations of these methods also exist). It is the latter that will be used in this chapter. A short introduction into transition path sampling (TPS) will be given below that closely follows Reference [141]. The reader is also referred to Reference [118].

6.1.1 The Transition Path Ensemble

The principle behind TPS is that there exist many paths that connect the reactant and product basins on the free energy surface. These are termed *transition paths*. Some of these transition paths will be preferred to others, in analogy to how some configurations are preferred to others in the canonical ensemble. In keeping with this analogy, we can therefore define the *transition path ensemble*, which is the weighted distribution of all possible transition paths. Specifically, consider a path $x(\mathcal{T})$ of length $\mathcal{T} = K\Delta t$ that has been discretised into an ordered sequence of states $x(\mathcal{T}) = \{x_0, x_{\Delta t}, x_{2\Delta t}, \ldots, x_{K\Delta t}\}$. The statistical weight

$P[x(\mathcal{T})]$ of observing any path (not necessarily a transition path) is given by:

$$P[x(\mathcal{T})] = \rho(x_0) \prod_{i=0}^{K-1} p(x_{i\Delta t} \rightarrow x_{(i+1)\Delta t}), \tag{6.1}$$

where $\rho(x_0)$ is the normalised distribution of initial phase space points (e.g. if $\{x_0\}$ is distributed according to the canonical ensemble, then $\rho(x_0) \propto \exp\{-\beta H(x)\}$) and $p(x_{i\Delta t} \rightarrow x_{(i+1)\Delta t})$ is the probability of the system evolving from state $x_{i\Delta t}$ to $x_{(i+1)\Delta t}$. The statistical weight of paths $P_{AB}[x(\mathcal{T})]$ that connect the reactant basin A to the product basin B can be written as:

$$P_{AB}[x(\mathcal{T})] \equiv \frac{h_A(x_0) P_{AB}[x(\mathcal{T})] h_B(x_{K\Delta t})}{\mathcal{Z}_{AB}(\mathcal{T})}, \tag{6.2}$$

where $h_A(x)$ is the population function of basin A, which is equal to 1 if x lays in A and is zero otherwise ($h_B(x)$ is defined analogously). The demoninator of Equation 6.2 is defined as:

$$\mathcal{Z}_{AB} \equiv \int \mathcal{D}x(\mathcal{T}) h_A(x_0) P[x(\mathcal{T})] h_B(x_{K\Delta t}). \tag{6.3}$$

The notation '$\int \mathcal{D}x(\mathcal{T})$' indicates summation over all pathways $x(\mathcal{T})$. By comparing to Equation 2.22, it is clear that $\mathcal{Z}_{AB}$ can be thought of as a restricted partition function in trajectory space. The aim of TPS is to sample paths according to Equation 6.2. This is achieved by a Monte Carlo procedure with a series of specially designed trial moves. These are briefly outlined below.

6.1.2 Monte Carlo Moves in Trajectory Space

In standard Monte Carlo, one samples configurations of the system by, for example, attempting small translational and rotational moves of the particles and accepting such moves with a probability $\exp(-\beta \Delta U)$, where ΔU is the change in potential energy upon performing such a trial move. In TPS, however, one is not interested in sampling configurations *per se*, but trajectories. It is therefore necessary to first establish ways to generate trial trajectories and then determine whether or not to accept such a 'move'. The two most common types of TPS move that satisfy detailed balance are the 'shooting' and 'shifting' moves

developed by Chandler and co-workers [118]. Even though these moves can be rigorously derived from classical statistical mechanics [141], the end result is that the generation and acceptance of these moves are straightforward and therefore only a description of the shooting and shifting moves are given.

6.1.2.1 Shooting

Given a transition path $x^{(\mathrm{o})}(\mathcal{T})$, a new path $x^{(\mathrm{n})}(\mathcal{T})$ can be generated by a shooting move according to the following procedure:

1. Select a time slice $x^{(\mathrm{o})}_{i\Delta t}$ (or 'shoot point') from $x^{(\mathrm{o})}(\mathcal{T})$ at random.
2. Modify the shoot point by adding a random displacement: $x^{(\mathrm{n})}_{i\Delta t} = x^{(\mathrm{o})}_{i\Delta t} + \delta x$. This random displacement should be consistent with the ensemble of initial conditions and be symmetric with respect to the reverse move. It should also obey any conservation laws (i.e. total linear and angular momentum).
3. Accept the new shoot point with probability $\min\left[1, \rho(x^{(\mathrm{n})}_{i\Delta t})/\rho(x^{(\mathrm{o})}_{i\Delta t})\right]$. If this is not accepted then go to step 6.
4. Integrate the equations of motion from the shoot point for a time $(K-i)\Delta t$ (i.e. until time $\mathcal{T}$ is reached). If $h_B(x^{(\mathrm{n})}_{K\Delta t}) \neq 1$ then go to step 6.
5. Integrate the equations of motion backwards in time from $x^{(\mathrm{n})}_{i\Delta t}$ for a time $i\Delta t$ (i.e. until time 0 is reached). If $h_A(x^{(\mathrm{n})}_{0\Delta t}) \neq 1$ then go to step 6.
6. If the $x^{(\mathrm{n})}(\mathcal{T})$ is rejected at any stage then retain $x^{(\mathrm{o})}(\mathcal{T})$. Otherwise, replace $x^{(\mathrm{o})}(\mathcal{T})$ with $x^{(\mathrm{n})}(\mathcal{T})$.

A few points are worth noting on the above algorithm. First, the random displacement performed at step 2 often involves changes only in the momenta. Second, in step 5, the backwards time integration is formally done with a negative time step (for time reversible integrators such as the Verlet algorithm - see Section 4.1). In practice, this is done by inverting the momenta at the shoot point, integrating the equations of motion forwards in time and then reversing all the momenta of the resulting backwards-facing segment of the trajectory. Finally, it should be noted that the above algorithm is for deterministic dynamics. The corresponding algorithm for stochastic dynamics is similar but with the exception that, due to the stochasticity, the random displacement at step 2 is not required and the dynamics can be propagated from the unmodified shoot point (after reversing momenta for

a backwards shoot). Furthermore, shooting need only be performed in one direction ('half-shooting'). This can be useful for diffusive barrier crossing processes such as nucleation [142].

6.1.2.2 Shifting

Again, given a transition path $x^{(o)}(\mathcal{T})$, a new path $x^{(n)}(\mathcal{T})$ can be generated by a shifting move according to the following procedure:

Forwards Shifting:

1. Randomly select a positive time interval $\delta t = m\Delta t$.
2. Copy the $K - m$ last time slices of the old path to the first $K - m$ time slices of the new path i.e. $x^{(n)}_{i\Delta t} = x^{(o)}_{(i+m)\Delta t}$ for $i = 0, \ldots, K - m$.
3. From $x^{(n)}_{m\Delta t}$, integrate the equations of motion forwards for a time $m\Delta t$.
4. Accept $x^{(n)}(\mathcal{T})$ if it is reactive and reject it otherwise.

Backwards Shifting:

1. Randomly select a positive time interval $\delta t = m\Delta t$.
2. Copy the $K - m$ first time slices from the old path to the $K - m$ last time slices of the new path i.e. $x^{(n)}_{(i+m)\Delta t} = x^{(o)}_{i\Delta t}$ for $i = 0, \ldots, K - m$.
3. From $x^{(n)}_{m\Delta t}$, integrate the equations of motion backwards for a time $m\Delta t$
4. Accept $x^{(n)}(\mathcal{T})$ if it is reactive and reject it otherwise.

Again it should be noted that the above algorithm is for deterministic dynamics and in practice, the backwards time integration is achieved by reversing the momenta at $x^{(n)}_{m\Delta t}$ and integrating forwards in time (and then reversing the momenta of the new trajectory segment at the end). This makes the algorithm applicable also to stochastic dynamics.

Equipped with these shooting and shifting moves, one then performs a TPS simulation much in the same way that regular Monte Carlo simulations are performed in configuration space. This includes allowing for relaxation from the initial trajectory ('equilibration') and running the TPS simulations for long enough such that enough of trajectory space has been sampled.

6.2 Introduction

In Chapter 2, the concept of the free energy as a function of a coordinate Q that separated the liquid and crystal basins was formally introduced (see Equations 2.25 and 2.26). Ordinarily, Q is a *structural order parameter* that distinguishes the liquid and the crystal based on the structural differences between these two states. Structure, however, is not the only way that a liquid can be distinguished from a crystal: it is an obvious statement that the dynamical properties of these two phases are also significantly different, with the liquid being a 'mobile' or 'ergodic' phase and the crystal being an 'immobile' or 'nonergodic' phase.[1] The work presented in this chapter treats the changes in dynamics between liquid water and ice on a more equal footing to the changes in structure. The motivation for such an approach is twofold. The first reason is practical. Even though many examples of using structural order parameters with free energy methods to simulate ice formation exist [74–76, 143–145], uncertainties remain regarding how well these structural order parameters approximate the actual reaction coordinate. Problems can arise because, although the structures of bulk liquid water and ice can be easily distinguished, at the interface between a crystal nucleus and liquid the situation is less clear-cut. This is further exacerbated in the case of ice formation, as the local structure of molecules in liquid water and ice are not too dissimilar. Investigating how the dynamics change during the phase transition therefore provides another potential tool for computational scientists to study ice nucleation. The second reason is more important and concerns the fundamental behaviour of supercooled liquids, including water. This will be discussed below.

6.2.1 The Dynamical Behaviour of Supercooled Liquids

The work presented in this chapter is inspired by and uses many concepts from the literature on the glass transition. Like a crystal, a glassy material is immobile and nonergodic, but unlike the crystal, lacks any associated long range order. Here, two phenomena highlighting the fundamental importance of particle dynamics in supercooled liquids are discussed, which act as motivation for explicitly

[1] This is not to say that dynamics alone can distinguish liquid and crystal states. By definition, a crystal has a periodically repeated microscopic structure.

considering changes in the dynamics of water during ice nucleation.

6.2.1.1 Phase Transitions in Trajectory Space

A major breakthrough in the theory of glass forming liquids occurred when Hedges *et al.* [146] demonstrated the existence of a nonequilibrium first order phase transition in an atomistic glass former. Here the key observations of Reference [146] will be briefly summarised.

An equilibrium phase transition, such as crystallisation, is controlled by traditional thermodynamic variables such as temperature and pressure. The nonequilibrium phase transition elucidated in Reference [146], however, is controlled by variables that drive the system out of equilibrium. Moreover, the nonequilibrium phase transition occurs in trajectory space rather than configuration space. To clarify these remarks, let us first consider an equilibrium phase transition in a system of N particles at a pressure p. The volume $V(\mathbf{r}^N) \equiv V$ can be used as an order parameter to describe the state of the system as different phases can be distinguished by the typical size of V. Changes in V are coupled to p; p is a thermodynamic field. More formally, the probability of observing a particular configuration $\mathbf{r}^N$ of the system is proportional to $P_0(\mathbf{r}^N) \exp\left[-\beta \Delta p V(\mathbf{r}^N)\right]$, where $P_0(\mathbf{r}^N)$ is the probability of $\mathbf{r}^N$ at some reference pressure $p_0 = p - \Delta p$. The mean volume of the system at a given pressure is denoted by $V_p \equiv \langle V(\mathbf{r}^N)\rangle_p$. A first order phase transition between phases "1" and "2" is manifested by a discontinuity at a pressure $p = p^*$. At p^*, phases 1 and 2 coexist with volumes per particle v_1 and v_2 respectively and the distribution of the volume $P_{p^*}(V)$ is bimodal, with a low probability of observing a value of V intermediate between Nv_1 and Nv_2. This low probability at intermediate values of V decreases exponentially with the free energy cost to form an interface between phases 1 and 2, which scales as $N^{1-1/d}$, where d is the dimensionality of the system. This is shown schematically in Figure 6.1. Furthermore, due to the presence of two macroscopic phases, at coexistence, the mean square fluctuations in the volume scales as N^2. As the response of the volume to changes in pressure is given by $-\partial V_p/\partial p = \beta\langle(V - V_p)^2\rangle_p$, the width of the crossover shown in Figure 6.1 decreases as $1/N$.

Let us now consider a nonequilibrium phase transition that occurs in trajectory space. First, a dynamical order parameter $\mathcal{K}[x(t)]$ needs to be defined. In

Figure 6.1: System size dependence of a first order phase transition. For an equilibrium first order phase transition, the volume changes discontinuously between volumes per particles v_2 and v_1 at the coexistence pressure p^*. In the case of the nonequilibrium phase transition, the mean dynamical activity $\mathcal{K}_s$ changes discontinuously between activities per particle per unit time k_a and k_i at the coexistence dynamical field $s = s^*$. At coexistence, the probability distributions $P_{p^*}(V)$ and $P_{s^*}(\mathcal{K})$ are bimodal, with configurations or trajectories at intermediate values being much higher in energy than those at the basins. Adapted from Reference [146].

Reference [146] the following was used:

$$\mathcal{K}[x(t)] = \Delta t \sum_{t=0}^{t_{\text{obs}}} \sum_{j=1}^{N} |\mathbf{r}_j(t + \Delta t) - \mathbf{r}_j(t)|^2, \tag{6.4}$$

where t_{obs} is the length of the trajectory and Δt is a time period over which a particle would typically move a distance on the order of a molecular diameter in the normal (i.e. not supercooled) liquid. It will be seen later that other dynamical order parameters can also be used. From the form of Equation 6.4, it is clear that $\mathcal{K}[x(t)]$ depends upon the system's history over the observation period $0 \leq t \leq t_{\text{obs}}$. When particles are mobile (like in the liquid), then $\mathcal{K}[x(t)]$ is large, whereas for immobile phases (such as a crystal or glass), $\mathcal{K}[x(t)]$ is small. Now consider the probability distribution of trajectories when $\mathcal{K}[x(t)]$ is coupled to a

field s, analogous to how the distribution of microstates was considered when V was coupled to the thermodynamic field p. This distribution is proportional to $P_0[x(t)] \exp\{-s\mathcal{K}[x(t)]\}$, where $P_0[x(t)]$ is the equilibrium distribution of trajectories (i.e. when $s = 0$). The partition function of the equilibrium distribution is trivial due to the constraint of normalisation $\sum_{x(t)} P_0[x(t)] = 1$. In the case of non-zero values of the field, however, the distribution:

$$P_s[x(t)] \propto P_0[x(t)] \exp\{-s\mathcal{K}[x(t)]\} \tag{6.5}$$

has a nontrivial partition function, which for $s > 0$ decreases with increasing t_{obs}. For large enough s, trajectories favoured by $P_s[x(t)]$ are those which are immobile/nonergodic. The distribution $P_s[x(t)]$ therefore describes a system that has been driven out of equilibrium. Through TPS simulations of $P_s[x(t)]$, Hedges *et al.* were able to establish the existence of a first order phase transition between an 'active' mobile liquid phase and an 'inactive' immobile glassy phase, manifested by a discontinuity in the mean value of the dynamical order parameter $\langle \mathcal{K}[x(t)] \rangle_s \equiv \mathcal{K}_s$ at a value of the field $s = s^*$. This is shown schematically in Figure 6.1.

The above discussion has introduced the concept of 'states' in trajectory space i.e. some paths are 'active' whereas other are 'inactive'. In the standard form of TPS introduced in Section 6.1, transitions paths that connect states in phase space are sampled. In Reference [146], however, the same techniques were also be used to estimate the density of states in trajectory space and the sampled paths were no longer reactive trajectories. A similar approach will be used in Section 6.3 where the dynamical and structural density of states in trajectory space will be considered.

6.2.1.2 Dynamic Heterogeneity

The above discussion of the existence of a dynamical phase transition is enlightening as it highlights the fundamental importance that the dynamics can have in defining the state of a system. It is, however, abstract in its formulation and the potential direct relevance to crystallisation is not obvious. The origin of this dynamical phase transition – *dynamic heterogeneity* – is, on the other hand, less

Figure 6.2: Dynamic heterogeneity in a two dimensional glass forming liquid. Reduced units specific to the system are used for length, time and temperature are used. In (a) the unprocessed coordinate of a tagged particle at time t is shown, along with the corresponding inherent structure and coarse grained coordinates (see Reference [147] for details). In (b) the displacement for each particle (in a 10 000 particle system) measured over a time interval Δt is shown for different temperatures. Blue indicates relative immobility, whereas red indicates relative mobility. At and below the onset temperature for glassy dynamics T_0, localised regions of mobility exist that are separated by regions of immobility. Taken from Reference [147].

abstract and can be more readily pictured in regular phase space.

In order to explain what is meant by 'dynamic heterogeneity', imagine measuring all of the positions of an N particle system at time t and then again at later time $t+\Delta t$ and calculating the single particle displacements $\{|\mathbf{r}_i(t+\Delta t)-\mathbf{r}_i(t)|\}$. It would be reasonable to expect a distribution of single particle displacements over this time period. In a system displaying dynamic heterogeneity, the distribution of single particle displacements is not spatially homogeneous over large length scales: the particles tend to form mobile and immobile domains. The extent of dynamic heterogeneity is temperature dependent, occurring after the temperature of the system falls below the system dependent onset temperature for glassy dynamics T_0. An example of a system displaying dynamic heterogeneity is shown in Figure 6.2 (b). At equilibrium, the mobile domains, or 'soft spots', diffuse through the system, combining and separating in the process. For systems that have fallen

out of equilibrium, such as in the glass, these soft spots are locked in place and do not evolve over experimental observation times.

With regards to water, both dynamic heterogeneity and dynamical phase transitions between the liquid and amorphous ices have been observed with computer simulations [148, 149]. Furthermore, the experimental onset temperature for glassy dynamics in water is $T_0 = T_{\max} \approx 277\,\mathrm{K}$ [150], where $T_{\max}$ is the temperature of maximum density of liquid water at ambient pressure. All ice nucleation in nature therefore occurs in systems displaying dynamic heterogeneity. Obvious questions thus arise such as "where does ice nucleation occur – in the mobile or immobile domains?" and "is nucleation a diffusion limited process?". To be able to address questions such as these, it is necessary to treat both the dynamics and structure of the system with equal importance. The results presented in this chapter have therefore been obtained with TPS, leaving the dynamics of the system unperturbed.

6.3 Free Energy Surfaces in Trajectory Space

Although the primary aim of this thesis is to understand how different surfaces affect ice formation, as the concepts (in the context of ice nucleation) are novel and implementation of the TPS algorithms was required, the scope of the current chapter is limited to homogeneous nucleation. Furthermore, a dynamical order parameter will not actually be used to drive ice formation. Rather, as a first step, ice formation will be driven using a structural order parameter and how the dynamics changes during crystallisation will be monitored.

To model bulk water, 216 mW molecules [80] under periodic boundary conditions have been used. Again it is noted that the melting temperature of mW water at ambient pressure is $274.6 \pm 1.0\,\mathrm{K}$, in excellent agreement with the experimental value of 273.15 K. Dynamics have been propagated using the velocity Verlet integration scheme as implemented in `LAMMPS` for the NPT ensemble [127], at a pressure of 1 atm. A time step of 5 fs was used for all simulations. To measure particle mobility, the indicator function [151]:

$$h_l(t) \equiv \Theta\left(|\bar{\mathbf{r}}_l(t + \Delta t) - \bar{\mathbf{r}}_l(t)| - a\right), \tag{6.6}$$

has been used, where $\Theta(x)$ is the heaviside step function. This indicator function has the property that $h_l(t) = 1$ if the l^{th} particle has moved further than a distance a in a time interval Δt. To distinguish non-trivial particle displacements from vibrations the inherent structure positions $\{\bar{\mathbf{r}}_l\}$ have been used (i.e. the positions obtained after energy minimisation using the `FIRE` optimisation algorithm [152], see Figure 6.2 (a)). Values of $\Delta t = 1\,\text{ps}$ and $a = 1\,\text{Å}$ were used, in accordance with previously established protocols [148, 151]. For a path $x(t) = \{x_0, x_{\Delta t}, \ldots, x_{K\Delta t}\}$ of length $t_{\text{obs}} = K\Delta t$ the dynamical order parameter is defined as:

$$c[x(t)] = \frac{1}{NK} \sum_{t=0}^{K-1} \sum_{l=1}^{N} h_l(t). \tag{6.7}$$

A structural order parameter for the path can similarly be defined:

$$Q_6[x(t)] = \frac{1}{K+1} \sum_{t=0}^{K} Q_6(t), \tag{6.8}$$

where $Q_6(t)$ is the sixth rank Steinhardt bond order parameter, defined as [153]:

$$Q_6(t) = \left(\frac{4\pi}{13} \sum_{m=-6}^{6} |\bar{Q}_{6m}(t)|^2 \right)^{1/2}, \tag{6.9}$$

where $\bar{Q}_{6m}(t)$ are defined as:

$$\bar{Q}_{6m}(t) = \frac{\sum_i^N \sum_j^{N_i^{\text{b}}(t)} Y_{6m}(\hat{\mathbf{r}}_{ij}(t))}{\sum_i^N N_i^{\text{b}}(t)}. \tag{6.10}$$

In Equation 6.10, $N_i^{\text{b}}(t)$ is the number of molecules within a distance of 3.5 Å of molecule i and $Y_{6m}(\hat{\mathbf{r}}_{ij}(t))$ are the sixth rank spherical harmonics that take the unit bond vector between molecules i and j as an argument. Note that, unlike the dynamical order parameter, the structural order parameter uses the true positions of the particles rather than the inherent structure positions.

In what follows, the 'shooting' and 'shifting' moves of transition path sampling (TPS) will be used to sample trajectory space. A 'massive stochastic thermostat' was applied by randomising the momenta of the particles according to the Maxwell-Boltzmann distribution at Δt intervals, following Reference [146]. To

sample paths of different values of $Q_6[x(t)]$ a bias is applied:

$$w[x(t)] = \frac{k_{Q_6}}{2}(Q_6[x(t)] - Q_6^{\min})^2, \tag{6.11}$$

where $Q_6^{\min}$ is a scalar quantity that defines the minimum of the bias potential and k_{Q_6} is a force constant that defines the steepness of the bias potential. One could apply a similar bias for the dynamical order parameter but for the moment, this has only been done for the structural order parameter. Along with the shooting and shifting TPS moves, replica exchange between systems with different values of k_{Q_6} and $Q_6^{\min}$ are also performed. This procedure amounts to umbrella sampling in trajectory space.

Unlike the glass transition, which is a non-equilibrium phenomenon [146], ice formation is governed by equilibrium statistical mechanics. As such, whereas long observation times are necessary to study the glass transition [151], preliminary tests suggested that short observation times are more convenient in this context, so that the system does not have time to transform from one state to another within the observation time. This allows one to unequivocally describe a path as being in the liquid or crystalline state. To this end, $t_{\text{obs}} = \Delta t = 1\,\text{ps}$ has been used in the simulations that follow. The free energy for the path ensemble is given by:

$$F(Q_6[x(t)]) = -\ln\{P(Q_6[x(t)])\}, \tag{6.12}$$

where $P(Q_6[x(t)])$ is the probability distribution of the structural order parameter. The free energy as a function of $c[x(t)]$ has an analogous definition. Similarly, from the joint probability distribution $P(Q_6[x(t)], c[x(t)])$ a two dimensional free energy can be defined. In Appendix F, this path free energy is related to the free energy of equilibrium statistical mechanics. To reconstruct the unbiased probability distribution functions from the umbrella sampling simulations, the weighted histogram analysis method [154, 155] was used.

In Figure 6.3 the free energy curves $F(Q_6[x(t)])$, $F(c[x(t)])$ and the free energy surface $F(Q_6[x(t)], c[x(t)])$ for bulk water at 220 K are shown. As expected, on the $F(Q_6[x(t)], c[x(t)])$ surface one sees a minimum corresponding to the liquid at low values of $Q_6[x(t)]$ and relatively high values of $c[x(t)]$. This minimum is shallower

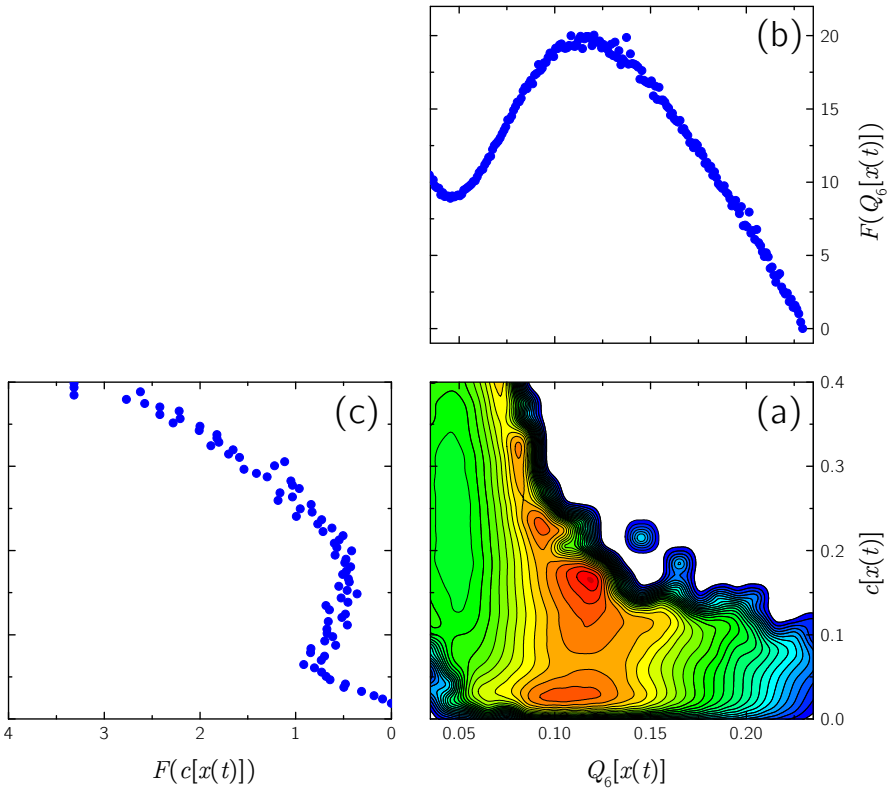

Figure 6.3: Path free energy surfaces for bulk mW water at 220 K. (a) The 2D free-energy surface as a function of $Q_6[x(t)]$ and $c[x(t)]$: the colour scale goes from blue to red, indicating low and high regions of free energy, respectively. (b) Path free energy as a function of $Q_6[x(t)]$ only. (c) Path free energy as a function of $c[x(t)]$ only: note that to correspond with (a), $c[x(t)]$ is plotted along the y-axis. From (a) one can see a distinct minimum at low values of $Q_6[x(t)]$ and relatively high values of $c[x(t)]$. This corresponds to the liquid state. At low values of $c[x(t)]$ there is a clear 'crystallisation channel' that spans the approximate range $0.07 < Q_6[x(t)] < 0.23$ and exhibits a barrier at $Q_6[x(t)] \approx 0.11$. This barrier is clearly seen in panel (b). Along this crystallisation channel, $c[x(t)]$ is more-or-less constant. From (c) it can be seen that crystallisation is essentially barrierless along $c[x(t)]$, with immobile states becoming preferable once the barrier in the $Q_6[x(t)]$ direction has been crossed.

along the $c[x(t)]$ direction than it is along the $Q_6[x(t)]$ direction. There is also a clear 'crystallisation channel' along which the system traverses from liquid to ice that exhibits a barrier at $Q_6[x(t)] \approx 0.11$. This occurs at low values of $c[x(t)]$. This barrier is clearly seen in the curve of $F(Q_6[x(t)])$. It can also be seen that the shallow part of the liquid minimum 'overlaps' with this crystallisation channel and that the value of $c[x(t)]$ remains roughly constant as the system becomes more structured. This is reflected in the plot of $F(c[x(t)])$, where it can be seen that the free energy profile is essentially barrierless to crystallisation along $c[x(t)]$.

Figure 6.4 shows the same data obtained at 260 K. Again a clear minimum

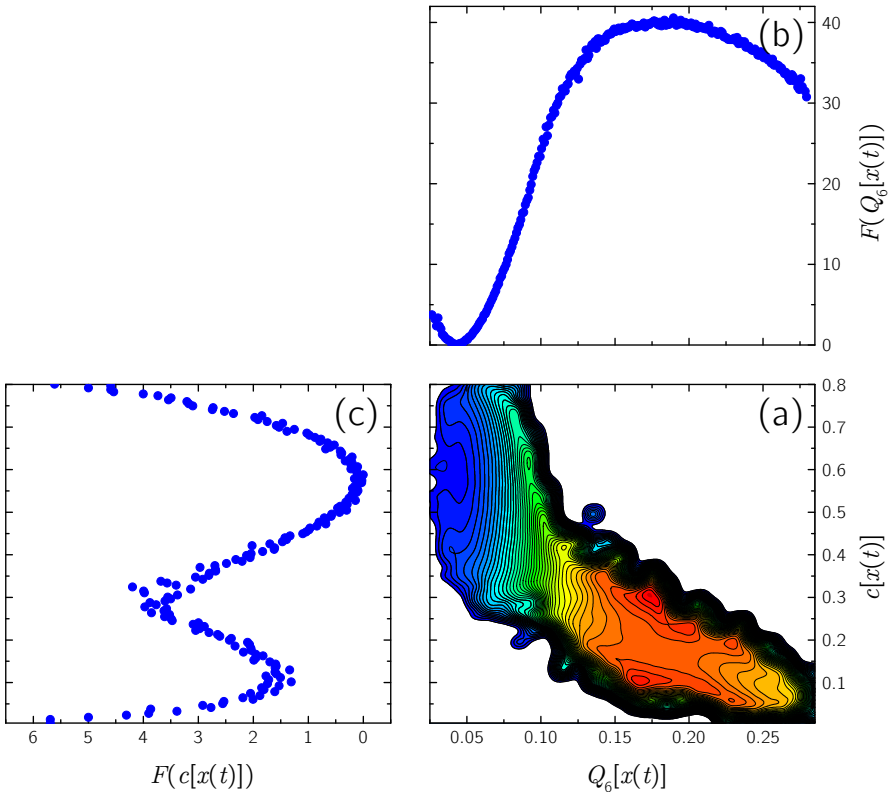

Figure 6.4: Path free energy surfaces for bulk mW water at 260 K. The panels are labelled as in Figure 6.3. Again in panel (a) a clear minimum corresponding to the liquid at low values of $Q_6[x(t)]$ and relatively high values of $c[x(t)]$ is seen. Unlike in Figure 6.3, however, the value of $c[x(t)]$ is decreasing along the crystallisation channel, which now exhibits a maximum at $Q_6[x(t)] \approx 0.18, c[x(t)] \approx 0.15$. From panel (c) it can now be seen that there is a barrier to crystallisation along $c[x(t)]$ separating the liquid and crystal states. This occurs at $c[x(t)] \approx 0.3$. The increase in $F(c[x(t)])$ at low values of $c[x(t)]$ is because sampling beyond $Q_6[x(t)] \approx 0.28$ has not been done; one expects the free-energy to keep falling as $c[x(t)]$ decreases.

corresponding to the liquid that is more shallow in the $c[x(t)]$ than $Q_6[x(t)]$ direction is seen. However, unlike at 220 K, $c[x(t)]$ does not remain constant along the crystallisation channel but is decreasing as the system becomes more structured. At this higher temperature, the structure and the dynamics are coupled. Furthermore, if one looks at $F(c[x(t)])$ a barrier at $c[x(t)] \approx 0.3$ is now seen. It is clear that the behaviour of the dynamics during crystallisation at these two temperatures is different.

The results from these path sampling simulations suggest that the dynamics at temperatures of 220 K and 260 K display different behaviours during bulk homogeneous ice nucleation. At 220 K, structuring occurs without any noticeable changes in the dynamics whereas at 260 K the dynamics and the structuring of the

system are coupled. A barrier to crystallisation in the dynamical order parameter at 260 K is also observed that is absent at 220 K. An initial interpretation of these results is that during crystallisation, at both temperatures, the mobility of the liquid first decreases before structuring occurs (inferred from the shallowness of the liquid basin along $c[x(t)]$).

6.4 TPS Applied to Ice Formation

To test the interpretation of the above results, the more regular style of TPS has been employed whereby reactive trajectories that connect the liquid and crystal basins are sampled. For each of these trajectories, the *instantaneous* values of the mobility and the structure have been measured. Initial analysis of these results generally support the conclusions drawn from the previous free-energy calculations.

As in the previous calculations, all simulations presented below are for bulk water, modelled using 216 coarse-grained mW [80] molecules under periodic boundary conditions. Simulations with 1728 mW molecules have also been performed and yield results consistent with those reported below (see Appendix F). Dynamics have been propagated using the velocity Verlet integration scheme as implemented in `LAMMPS` for Langevin dynamics and the NPH ensemble [127], at a pressure of 1 atm. For the Langevin dynamics, a large damping time of 10 ps was chosen so that the bulk diffusion coefficient of mW water obtained in the NPT ensemble was recovered.[2] A time step of 5 fs was used for all simulations.

TPS simulations were performed using the shooting and shifting moves appropriate for stochastic dynamics. As stochastic dynamics is being used, the fact that 'half-shooting' moves i.e. shooting only either forwards or backwards in time, also generate transition paths of significant dynamical weight (which is not the case for deterministic dynamics) is exploited; this has been shown to dramatically increase the efficiency of TPS for diffusive barrier crossing [142]. The sampled paths were $\mathcal{T} = 200$ ps in length and were defined as reactive if $Q_6[x(0)] < 0.07$

[2]For a system of 1728 bulk mW molecules at 298 K and 1 bar, the self-diffusion coefficient was calculated to be $6.2 \pm 0.3 \times 10^{-5}$ cm^2 and $6.3 \pm 0.2 \times 10^{-5}$ cm^2 s^{-1} in the NPT and NPH ensembles, respectively. With the Langevin dynamics used in these simulations, a value of $6.3 \pm 0.1 \times 10^{-5}$ cm^2 s^{-1} was obtained.

and $Q_6[x(\mathcal{T})] > 0.30$. Trial paths were accepted if they were deemed reactive, otherwise they were rejected. Shooting and shifting moves were attempted in a ratio of 3:2. The maximum shifting length was 30 ps. Positions and momenta were stored at 1 ps intervals.[3]

Although the Steinhardt Q_6 structural order parameter is being used to define the liquid and crystal basins, it is also useful to monitor the crystallinity of individual water molecules based upon their local environment throughout the trajectory. To this end, the rank-3 local Steinhardt bond order parameter q_3 [74, 111] has been employed. Specifically, the 7-dimensional local bond order vector $\mathbf{q}_3(i)$ around a water molecule i is defined as:

$$q_{3m}(i) = \frac{1}{N_i^{\mathrm{b}}} \sum_{j=1}^{N_i^{\mathrm{b}}} Y_{lm}(\hat{\mathbf{r}}_{ij}), \tag{6.13}$$

where N_i^{b} is the number of water molecules within 3.5 Å of molecule i, and $\hat{\mathbf{r}}_{ij}$ is the unit bond vector between molecules i and j. One can then calculate $d_3(i,j) = \hat{\mathbf{q}}_3(i) \cdot \hat{\mathbf{q}}_3(j)$, from which

$$n_{\mathrm{c}}(i) = \sum_{j=1}^{N_i^{\mathrm{b}}} \Gamma(d_3(i,j)) \tag{6.14}$$

is computed, where

$$\Gamma(x) = \begin{cases} 1, & \text{if } (x < -0.80) \wedge (-0.20 < x < -0.05); \\ 0, & \text{otherwise.} \end{cases} \tag{6.15}$$

Finally, molecule i is defined as ice-like if $n_c(i) \geq 3$. As well as the local structure of each molecule, how the mobility of each particle varies over the course of the trajectory is also measured using the procedure outlined in Section 6.3.

Four TPS simulations were performed at both 220 K and 260 K, each with

[3] As Langevin dynamics is used, it is not necessary to change the momenta at the shoot point. Rigorously, the state of the barostat should also be taken into account, but this is technically challenging to implement. As stochastic dynamics are being used, it might be possible to justify ignoring the state of the barostat, but a rigorous justification is work in progress. For the moment, the results seem reasonable (and agree with the previous free energy simulations), so it is assumed that any effects of neglecting the barostat are insignificant.

a different initial trajectory. At 220 K, one of these initial trajectories was taken from a 200 ps TPS+US simulation that had spontaneously nucleated: two other initial trajectories were generated from this initial trajectory by performing short TPS simulations (these three initial trajectories are referred to as 'TPS+US trajectories'). The remaining initial trajectory was generated by reversing a melting trajectory at 370 K and rescaling each velocity component by $\sqrt{\frac{220}{370}}$ (this is referred to as the 'reverse-melt trajectory'). By using a completely independent initial trajectory, one should be able to gauge the robustness of the results. For the simulations at 260 K the same initial trajectories were used with the velocity components rescaled by $\sqrt{\frac{260}{220}}$. For each simulation, between 13800 and 29400 TPS moves were performed, with trajectories stored every 100 TPS moves. Approximately 57% and 64% of TPS moves were accepted at 220 K and 260 K, respectively.

In Figure 6.5 snapshots along a typical trajectory obtained at 220 K are shown, with mobile particles depicted as solid red spheres and immobile particles as transparent blue spheres. It can be seen that even in the initial liquid state, the system consists of predominantly immobile molecules with localised areas of mobility ('soft-spots'). As the trajectory proceeds one sees that ice begins to form within a relatively immobile region. The snapshots from the simulation at 260 K (Figure 6.6) provide a qualitatively different picture. First, the initial system is much more mobile than at 220 K. Second, as the simulation proceeds, one can see a clear decrease in the mobility, with localised regions of immobile particles ('hard-spots') appearing without any significant indications of ice formation. Similar to the simulation at 220 K, however, is that when ice does form, it does so within one of these localised hard-spots.

In Figure 6.7 the time evolution of the number of mobile and ice-like molecules is shown for the trajectories presented in Figure 6.5 and 6.6. As expected, at 220 K, a drastic change is not seen in the number of mobile particles over the course of the trajectory, although there is a notable decrease in the size of the fluctuations as the ice grows. At 260 K, however, the growth of ice does appear to be preceded by a drop in the mobility (at a time approximately between 60-90 ps) and as the ice grows, the mobility is clearly decreasing.

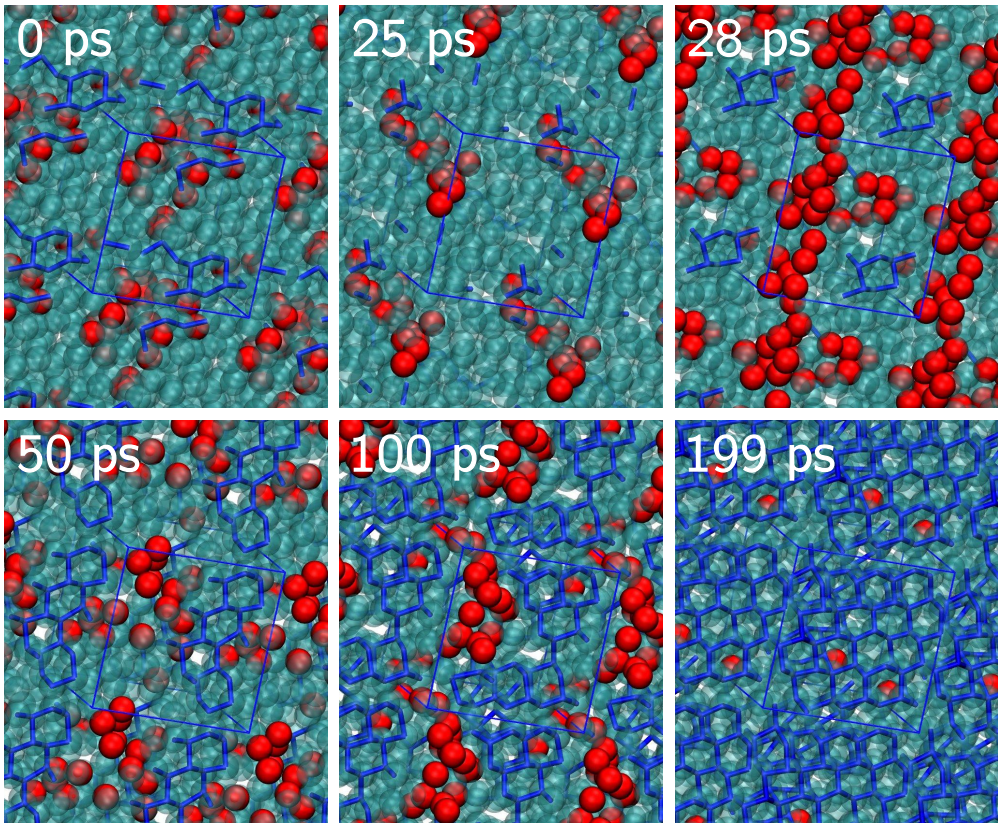

Figure 6.5: Snapshots along a typical trajectory at 220 K. Solid red spheres and transparent blue spheres indicate mobile and immobile particles, respectively. Bonds between ice-like particles are shown by solid blue lines. The time along the trajectory is indicated in the top left corner of each panel. One can clearly see that the system is mostly immobile throughout the trajectory, and that ice forms within a relatively immobile region.

Overall, the results from these two simulations appear to support the interpretation of the results from Section 6.3, with the exception that we do not see a decrease in mobility prior to ice formation at 220 K. In retrospect this is not unreasonable because, as previously noted, the liquid basin overlaps with the crystallisation channel. In other words, although one may expect the mobility to be lower along the crystallisation channel than the average value observed in the liquid basin, it will not be noticeably lower than low-mobility amorphous states accessible through thermal fluctuations.

Of course, one should not rely upon the results of two trajectories alone, and should look to all of the simulations harvested from the transition path ensemble with TPS. In order to do this, the following algorithm has been devised:

1. Starting from the initial frame, scan through the trajectory. If the

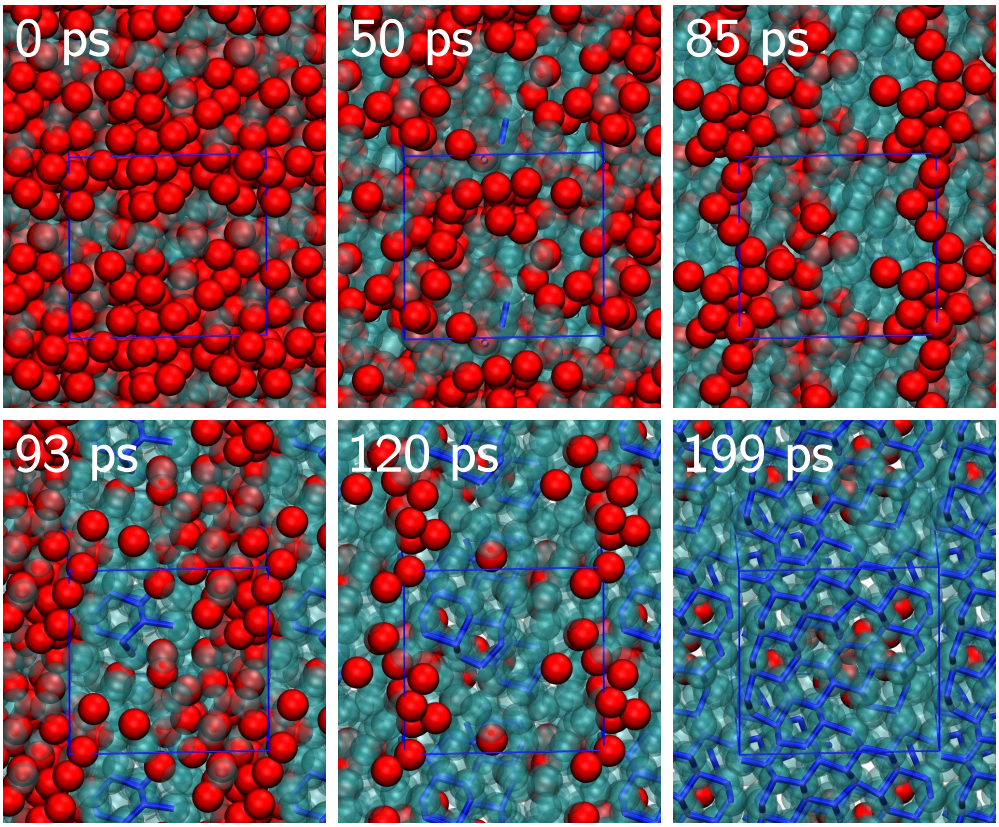

Figure 6.6: Snapshots along a typical trajectory at 260 K. The colour code is the same as in Figure 6.5. As expected, mobility is initially much higher at 260 K than 220 K. It can also be seen that ice initially forms in a region of space that is relatively immobile.

number of ice-like molecules is greater than or equal to $N_{\text{threshold}}$ then stop, and set time = 0 at this frame.

2. From time = 0, loop backwards over the frames, storing the number of mobile/ice-like molecules at each frame. Each frame is labelled $-i\Delta t$, where i is the offset from time = 0 and Δt is the time interval between frames.
3. Repeat step 2 in the forward direction (labelling frames $+i\Delta t$).
4. Load a new trajectory and return to step 1.
5. Once all trajectories have been analysed, take the average number of mobile/ice-like molecules at each frame.

In Figures 6.8 and 6.9 the results of this analysis are shown with $N_{\text{threshold}} = 50$ for trajectories harvested at 220 K and 260 K, respectively.[4] The shaded regions around the lines indicate a 95% confidence interval. At 220 K it appears that the mobility steadily drops over the course of the simulation and much more gradually than the rate at which ice grows. Even though the mobility is steadily decreasing,

[4]Values of $N_{\text{threshold}} = 25$ and 75 were also tested and yielded qualitatively similar results.

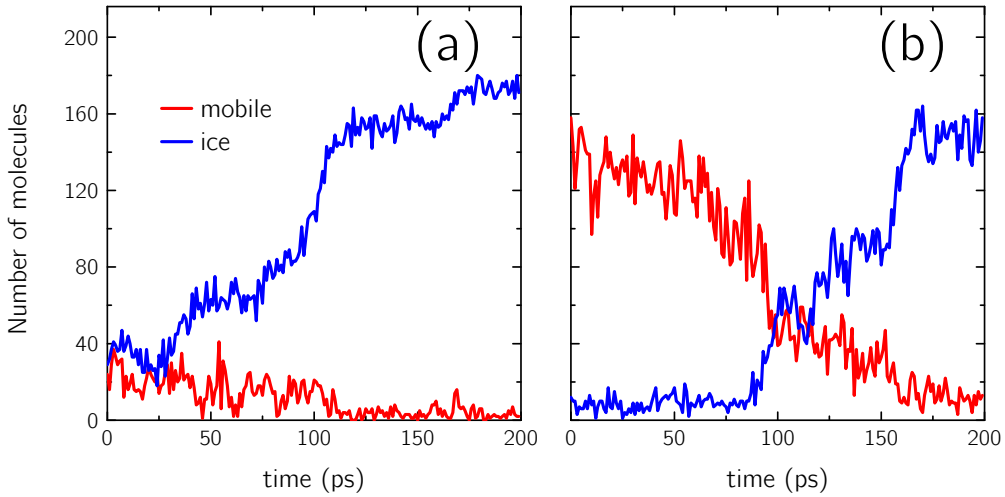

Figure 6.7: Time evolution of the number of mobile and ice-like particles for a single trajectory. (a) The simulation at 220 K shown in Figure 6.5. (b) The simulation at 260 K shown in Figure 6.6. At 260 K there appears to be a decrease in mobility between *ca.* 60-90 ps that precedes the growth of ice, and as ice grows, the mobility is clearly decreasing. At 220 K, there is no clear decrease in mobility before ice growth, and there is also far less of a decrease in mobility as ice grows than at 260 K.

one does not see a significant drop in the number of mobile particles prior to nucleation. There is, however, certainly no indication of enhanced mobility facilitating nucleation at this temperature [156, 157]. Turning our attention to the results at 260 K, a significant drop in the mobility prior to nucleation is now seen. In all four TPS simulations, one sees that ice begins to grow between -20 and -10 ps, but that the mobility begins to decrease as early as -40 ps. Furthermore, the decrease in mobility is much faster than at 220 K and appears comparable to the rate of ice growth.

In the above analyses, the structural order parameters used have been sensitive to the formation of the 'regular' cubic and hexagonal ice structures and the observed changes in mobility could therefore be associated with the formation of other structures that are 'invisible' to the Q_6 and q_3 order parameters. In order to test for other structures, a local order parameter similar to q_3 but based on the 12^{th} rank spherical harmonics has also been used (denoted q_{12}). This order parameter (the details of which are given in Appendix F) can detect not only cubic and hexagonal ice-like molecules, but also molecules belonging to the clathrate hydrate crystal structures, and 'ice 0', which is a recently proposed metastable form of ice. It is interesting to consider ice 0 because its recent *in silico* discovery

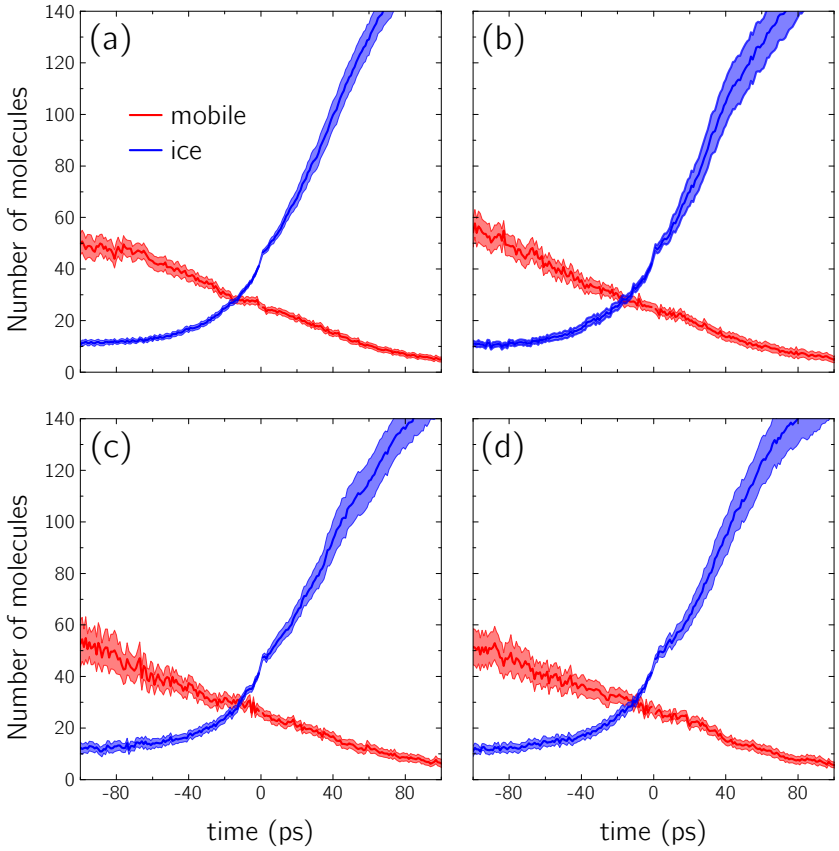

Figure 6.8: Time evolution of the number of mobile and ice-like particles at 220 K averaged over the transition path ensemble. Panels (a), (b) and (c) show results using the TPS+US trajectories as the initial trajectories and (d) shows the result of using the reverse-melt trajectory as the initial trajectory. These results have been obtained with $N_{\text{threshold}} = 50$ (see text). Although there is no pronounced drop in mobility prior to ice growth, it appears that the mobility drops steadily throughout the trajectory, and much more gradually than ice grows.

was also accompanied by a claim that it plays an important role in homogeneous ice nucleation *via* an Otswald's ripening mechanism [158]. The results of this analysis are presented in Figure 6.10 for trajectories harvested at both 220 K and 260 K (for clarity, only data derived from one of the initial trajectories is presented, but all TPS simulations give similar results).[5] It is clear that the increase in the number of crystalline molecules detected by q_{12} does not precede those detected by q_3. As q_{12} is sensitive to a comprehensive range of crystal structures, it is concluded that the decrease in immobile particles is not associated with any noticeable change in structure. This result also suggests that the importance of

[5]Due to the cost of this analysis, only a subset of trajectories have been sampled (approx. 50 trajectories, equally spaced throughout the TPS simulation in each case).

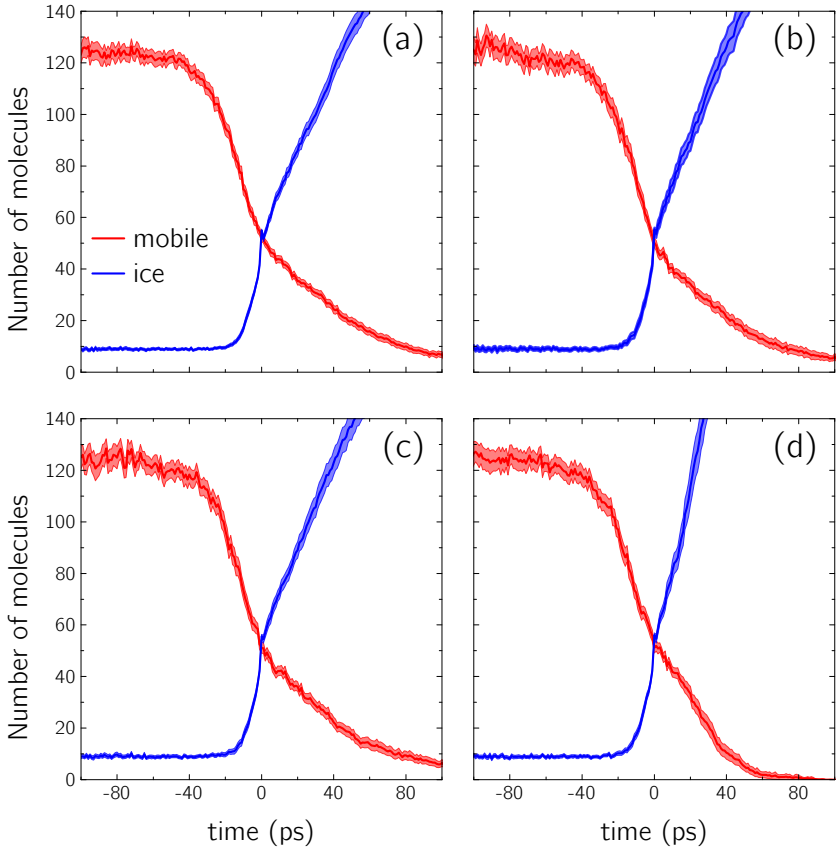

Figure 6.9: Time evolution of the number of mobile and ice-like particles at 260 K averaged over the transition path ensemble. Panels (a), (b) and (c) show results using the TPS+US trajectories as the initial trajectories and (d) shows the result of using the reverse-melt trajectory as the initial trajectory. These results have been obtained with $N_{\text{threshold}} = 50$ (see text). At *ca.* −40 ps one sees that the mobility begins to drop and that ice does not start to grow until *ca.* −15 ps. The mobility also decreases rapidly with ice growth.

ice 0 in homogeneous nucleation needs clarifying.

6.5 Summary

TPS simulations have been used to investigate how the dynamics of water changes during ice nucleation at 220 K and 260 K. First, by sampling short paths whose degree of crystallinity is well defined, the path free energy profiles projected onto both structural and dynamical order parameters were constructed. Then, using TPS to sample transition paths that connect the liquid and crystal basins, the actual time evolution of mobility and structure during ice nucleation have been measured. Both sets of simulations yield consistent results.

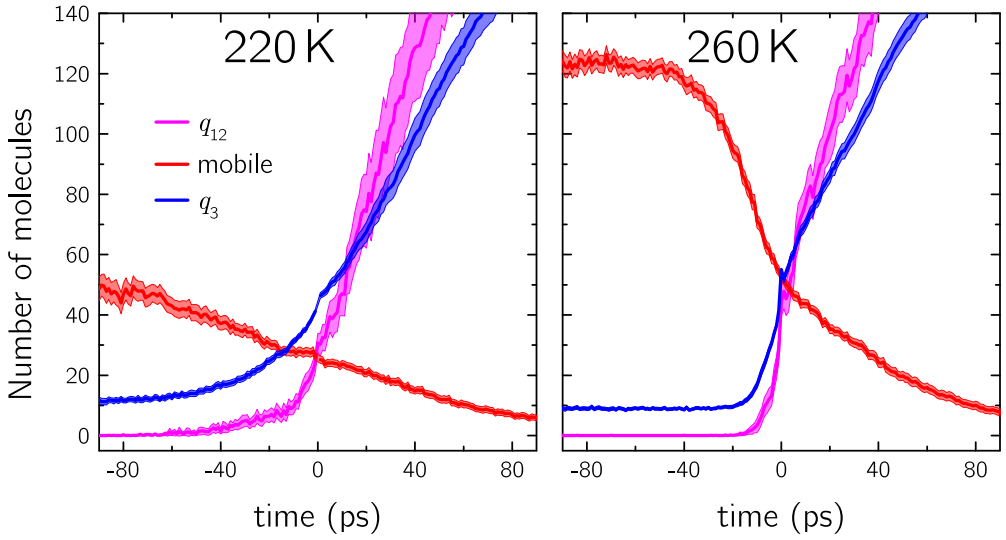

Figure 6.10: Time evolution of other crystal structures. The red and blue data are the same as in Figures 6.8 (a) and 6.9 (a) for 220 K and 260 K respectively. The magenta lines show the number of crystal-like molecules using the q_{12} order parameter, which is sensitive not only to cubic and hexagonal ice structures, but also clathrate hydrate structures and ice 0. There is no indication that ice 0 plays an important role in homogeneous nucleation.

At both temperatures, it is found that ice nucleation occurs in immobile domains or 'hard-spots'. At the lower temperature of 220 K, the hard-spots dominate and consequently, no significant drop in mobility precedes ice formation. On the other hand, at 260 K a clear drop in mobility is observed before any structuring occurs. This is true not only for cubic and hexagonal ice-like molecules, but also for the newly found metastable phase, ice 0, suggesting that this decrease in mobility is not associated with the formation of any particular crystal-like structure.

Given these results, it is clear that the 'dynamical state' of the system plays an important role in the nucleation of ice. As discussed in Section 2.2, classical nucleation theory concerns itself predominantly with the free energy to form a structured nucleus within the metastable stable liquid, with the dynamics only entering through the kinetic prefactor term in the expression for the rate (see Equation 2.32). Thus even at the qualitative conceptual level, CNT is lacking in its description of ice nucleation. Of particular concern is the picture of the crystal growing molecule-by-molecule [15]. Dynamic heterogeneity exists due to the presence of correlated, collective molecular motions in supercooled liquids: exactly how these collective molecular motions quantitatively affect the rate of

ice formation remains an open question.

Although homogeneous nucleation has been the focus of this chapter, it is interesting to discuss the possible consequences these findings might have for heterogeneous ice nucleation. Given that ice formation occurs in hard-spots in the liquid, the domain structure that arises to dynamic heterogeneity will be an influential factor in ice nucleation. Understanding how particles affect not only the structure of interfacial water but also the macroscopic domain structure of the dynamics may therefore also be important. It is not clear that such effects would necessarily act to promote ice nucleation.

Chapter 7

Methane Hydrate: evaluating the performance of density functional theory against diffusion Monte Carlo

The diffusion Monte Carlo results presented in this chapter were performed by Dr. Michael Towler and Prof. Dario Alfè at University College London.

7.1 Density Functional Theory

So far in this thesis, empirical force fields have mainly been used (whether all-atom or coarse grained) to define the total potential energy of the system. The reason for this is that these methods offer a computationally tractable way of exploring configuration space (see Figure 2.1). In certain cases, such as when covalent bonds are broken or under conditions unfamiliar to the parametrisation of the force field, an approach is needed that explicitly takes the electronic structure into account. The most commonly used [159] electronic stucture theory in condensed phase materials science is *density functional theory* (DFT). The basic concepts of DFT will be outlined below. A more thorough treatment can be found in e.g. References [160] and [161]. Throughout this section, atomic units are used such that $e^2 = \hbar = m = 1$, where e and m are the charge and mass of an electron,

respectively, and $\hbar$ is the reduced Planck's constant.

In the Born-Oppenheimer approximation the total Hamiltonian is:

$$\hat{H} = \hat{T}^{\text{e}} + \hat{V}^{\text{nuc-nuc}} + \hat{V}^{\text{e-nuc}} + \hat{V}^{\text{e-e}}. \tag{7.1}$$

The first term $\hat{T}^{\text{e}}$ is the total kinetic energy of the N electrons:

$$\hat{T}^{\text{e}} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2. \tag{7.2}$$

where ∇_i^2 is the Laplacian operator acting on each of the i electrons. The last three terms in Equation 7.1 are operators that give the nuclear-nuclear, electron-nuclear and electron-electron potential energies, respectively. For a system with M nuclei, these can be written as:

$$\hat{V}^{\text{nuc-nuc}} = \frac{1}{2}\sum_{I}^{M}\sum_{J\neq I}^{M}\frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}; \tag{7.3}$$

$$\hat{V}^{\text{e-nuc}} = -\sum_{i}^{N}\sum_{I}^{M}\frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|}; \tag{7.4}$$

$$\hat{V}^{\text{e-e}} = \frac{1}{2}\sum_{i}^{N}\sum_{i\neq j}^{N}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}. \tag{7.5}$$

In the above, Z_I is the charge of nucleus I at position $\mathbf{R}_I$ and $\mathbf{r}_i$ is the position of electron i. To get the total energy $E(\mathbf{R}^M)$ for a particular nuclear configuration $\mathbf{R}^M$ one has to solve Schrödinger's equation:

$$\begin{aligned}\hat{H}\psi^{\text{e}}(\mathbf{r}^N;\mathbf{R}^M) &= (\hat{T}^{\text{e}} + \hat{V}^{\text{nuc-nuc}} + \hat{V}^{\text{e-nuc}} + \hat{V}^{\text{e-e}})\psi^{\text{e}}(\mathbf{r}^N;\mathbf{R}^M) \\ &= E(\mathbf{R}^M)\psi^{\text{e}}(\mathbf{r}^N;\mathbf{R}^M)\end{aligned} \tag{7.6}$$

where the electronic wave function $\psi^{\text{e}}(\mathbf{r}^N;\mathbf{R}^M)$ depends parametrically on the nuclear coordinates. For all but the simplest systems, however, there are no analytical solutions to Equation 7.6. The biggest problem is $\hat{V}^{\text{e-e}}$, which couples the $3N$ spatial coordinates of the electrons. The defining characteristic of DFT

is that it aims to determine the ground-state one-electron density $n_0(\mathbf{r})$[1] of the system instead of the many-body wave function $\psi^{\mathrm{e}}(\mathbf{r}^N)$ itself, thus reducing the dimensionality of the problem to 3 rather than $3N$.

As a practical method, DFT became useful after the Hohenberg-Kohn (HK) theorems [162]. The first HK theorem (which will be stated without proof) shows that the ground state electronic wave function ψ_0^{e} is a unique *functional* of the ground-state one-electron density:

$$\psi_0^{\mathrm{e}} = \psi^{\mathrm{e}}[n_0]. \tag{7.7}$$

It therefore follows that all observables are functionals of the ground-state density. In particular, the ground-state electronic energy E_0^{e} is a functional of n_0 with one-to-one correspondence:

$$E_0^{\mathrm{e}} = E^{\mathrm{e}}[n_0] = \min_{n(r)} \langle \psi^{\mathrm{e}}[n] | \, \hat{T}^{\mathrm{e}} + \hat{V}^{\text{nuc-nuc}} + \hat{V}^{\text{e-nuc}} + \hat{V}^{\text{e-e}} \, | \psi^{\mathrm{e}}[n] \rangle \tag{7.8}$$

(a second HK theorem exists that states that $E_0^{\mathrm{e}} = E^{\mathrm{e}}[n_0] \leq E^{\mathrm{e}}[n]$, where the equivalence holds iff $n = n_0$). The interaction between the electrons and the nuclei can be written explicitly in terms of $n(\mathbf{r})$:

$$\begin{aligned} V^{\text{e-nuc}}[n] &= \int \mathrm{d}\mathbf{r} \sum_I^M \frac{n(\mathbf{r}) Z_I}{|\mathbf{r} - \mathbf{R}_I|} \\ &= \int \mathrm{d}\mathbf{r} \, V(\mathbf{r}) n(\mathbf{r}) \end{aligned} \tag{7.9}$$

The most difficult part of Equation 7.8 is an accurate evaluation of the kinetic energy functional: direct evaluation is currently infeasible without introducing significant errors. The most widely used method to obtain $E^{\mathrm{e}}[n]$ is that of Kohn and Sham [163], where one writes:

$$E^{\mathrm{e}}[n] = T_s^{\mathrm{e}}[n] + V^{\text{e-nuc}}[n] + V^{\text{Hartree}}[n] + E^{xc}[n] \tag{7.10}$$

where $T_s^{\mathrm{e}}[n]$ is the kinetic energy functional of a *non-interacting* reference system

[1] The one-electron density is defined such that $N = \int \mathrm{d}\mathbf{r} \, n(\mathbf{r})$.

of electrons, $V^{\text{Hartree}}[n]$ is the classical electron-electron Coulomb repulsion:

$$V^{\text{Hartree}}[n] = \frac{1}{2}\iint \mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}'\,\frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}, \tag{7.11}$$

and the exchange-correlation functional $E^{xc}[n]$ contains all the complicated quantum mechanical and many-body effects:

$$E^{xc}[n] = \left[(T^{\text{e}}[n] - T_s^{\text{e}}[n]) + (V^{\text{e-e}}[n] - V^{\text{Hartree}}[n])\right]. \tag{7.12}$$

One can then define an effective potential:

$$\begin{aligned} v^{\text{eff}}(\mathbf{r}) &= \frac{\delta\left\{V^{\text{e-nuc}}[n] + V^{\text{Hartree}}[n] + E^{xc}[n]\right\}}{\delta n(\mathbf{r})} \\ &= V(\mathbf{r}) + \int \mathrm{d}\mathbf{r}'\,\frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + \frac{\delta E^{xc}[n]}{\delta n(\mathbf{r})}, \end{aligned} \tag{7.13}$$

from which one obtains the single particle Schrödinger equation:

$$\left\{-\frac{1}{2}\nabla_i^2 + v^{\text{eff}}(\mathbf{r})\right\}\phi_{0,i}(\mathbf{r}) = \epsilon_{0,i}\phi_{0,i}(\mathbf{r}). \tag{7.14}$$

The Kohn-Sham orbitals $\{\phi_{0,i}(\mathbf{r})\}$ reproduce the one-electron density of the many-body system:

$$n(\mathbf{r}) = \sum_{i=1}^{N} |\phi_{0,i}(\mathbf{r})|^2. \tag{7.15}$$

As $n(\mathbf{r})$ depend on $\{\phi_{0,i}(\mathbf{r})\}$, Equation 7.14 has to be solved by an iterative procedure until self-consistency is reached.

In Kohn-Sham DFT, all of the approximations are made in $E^{xc}[n]$. Although $E^{xc}[n]$ only contributes a relatively small amount to the total electronic energy of the system, it is crucial for an accurate description of the interactions between atoms and molecules. Many approximate forms of $E^{xc}[n]$ exist, some of which will be used below to describe a methane water system. It will become apparent just how dependent intermolecular interactions can be on $E^{xc}[n]$.

7.2 Introduction

The focus of this thesis has been to understand heterogeneous ice nucleation, but a significant amount of effort has also been invested in trying to understand the formation of natural gas hydrates. Although the majority of these results are not presented, we finish by looking at a piece of work where the performance of currently used methods to describe the potential energy surface of methane hydrate is evaluated.

The clathrate hydrates of natural gases - crystalline compounds in which gas is dissolved in a host framework of water molecules - are important to a wide variety of applications across the energy and climate sciences. For example, the fact that one volume of hydrate can generate up to 180 volumes of gas upon dissociation at standard temperature and pressure, whilst only 15% of the recovered energy is required for dissociation, means that hydrate reservoirs are a potential untapped energy resource [3]. Even though there remains uncertainty in the total amount of hydrated gas on Earth, there is a consensus that this amount exceeds conventional gas reserves by at least an order of magnitude [164]. Perhaps a more pressing issue is that hydrates also pose a severe problem for flow assurance in oil and gas pipelines: if the mixed phases of water and natural gas are allowed to cool, hydrates may form and block the line, causing production to stall. As readily available oil and gas reserves become depleted, and the need for extraction from deeper reservoirs increases, the consequences of hydrate formation are becoming more severe. Although chemicals for inhibiting hydrate formation exist, they have generally been found on a trial-and-error basis, with little understanding of how they work at the molecular scale. This state of affairs has arisen from the fact that we have little knowledge of the fundamental mechanisms that underlie hydrate formation. Consequently, computer simulation has been used in recent years in attempts to improve our molecular level understanding of hydrate formation [165–169]. It is important, therefore, to understand both the molecular interactions present in condensed phase gas hydrates, and the performance of current approximations used to describe these interactions.

By far the most commonly used electronic structure method for investigat-

ing condensed phase systems is DFT (a recent review of DFT and the current challenges it faces is given in Reference [159]). Despite incredible success in its application to a wide variety of systems, DFT has a number of limitations. Of particular relevance to gas hydrates is the known deficiency of the local density approximation (LDA) and generalized gradient approximation (GGA) varieties of exchange-correlation (xc) functionals to account for van der Waals (vdW) dispersion interactions. Incorporating an accurate description of vdW interactions into density functional theory is a very active research area, with recent developments including Grimme's dispersion correction methods [170, 171], the Tkatchenko-Scheffler scheme [172] and the fully self-consistent vdW-DF method of Dion *et al.* [173] and its various derivatives [58, 59]. For a recent overview of these and other methods to incorporate vdW interactions into DFT see Reference [174]. Understanding the contribution of vdW to the bonding in solids is an important issue and there is a need to understand the strengths and weaknesses of various vdW-inclusive methods in order to improve the performance of DFT. Recent work has shown that vdW-inclusive DFT methods offer a systematic improvement over GGA functionals in describing the phase behaviour of ice [175, 176]. Like ice, gas hydrates also have an extended hydrogen bonded network of water molecules, but unlike ice, they contain cavities that gas molecules can occupy. Natural gas hydrates therefore offer the opportunity to test the ability of vdW-inclusive methods to simultaneously describe both the hydrogen bonded water network and the predominantly dispersion bound water-gas interaction. As well as DFT, force fields (FFs) are often used to investigate gas hydrates [165–167, 177, 178], especially when long time and length scales are required, such as in the study of nucleation processes.

Evaluating the performance of techniques such as DFT or FFs requires high-quality reference data to compare to – something that is lacking for gas hydrates in the condensed phase. For example, previous DFT studies [179, 180] have evaluated the performance of the chosen xc functional through comparison to experiment or quantum chemical methods on isolated clusters. However, various issues can arise when validating the performance of DFT to experiment, such as temperature/pressure, non-stoichiometry and quantum nuclear effects. Further-

more, although a source of valuable information for understanding the nature of interactions, comparison to isolated clusters (to which accurate quantum chemical methods are generally limited to) does not directly tell us how DFT methods are performing for the condensed phase. The tendency to validate FFs used in molecular dynamics or Monte Carlo simulation against experiment is even greater than it is for DFT. One method that has been shown to provide accurate energies for condensed phase water systems is diffusion Monte Carlo (DMC). DMC can be applied to a range of systems, both isolated and periodic [181, 182], has mild scaling behaviour [183, 184] and has rapid and automatic basis-set convergence [185]. DMC has also been shown to favour well in comparison to CCSD(T) – the so-called 'gold-standard' quantum chemical method – for calculations on the water dimer and other small water clusters [186–189]. It also gives a good description of the relative energies of different ice phases [175, 176] and has recently been shown to achieve sub-chemical accuracy for non-covalent interactions in the gas phase [190]. One can therefore have confidence that DMC can be used to obtain accurate reference data for periodic gas hydrate crystals. Specifically, DMC will be used to calculate accurate data for the energetics of a methane hydrate crystal.

In this chapter, the performance of a number of different *xc* functionals will be compared, ranging from the LDA and PBE [113] levels of approximation, in which vdW interactions are not accounted for, to a variety of dispersion-corrected functionals, namely: an empirical correction scheme from Grimme (PBE-D2) [170]; the method developed by Tkatchenko and Scheffler (PBE-vdW$^{\mathrm{TS}}$) [172], which like PBE-D2 involves an explicit summation of pairwise vdW dispersion interactions over all atom pairs, but differs in that the vdW C_6 coefficients are themselves functionals of the electron density; and a number of functionals from the vdW-DF family. In particular, the original vdW-DF of Dion *et al.* and the modified versions of Klimeš *et al.* [58, 59], in which the exchange functional is changed from that of revPBE, to 'optPBE', 'optB88' and 'optB86b' will be considered. These modified versions of vdW-DF have been shown to offer good performance for a wide range of systems [58, 59, 191, 192]. Throughout the rest of the chapter, the original vdW-DF of Dion *et al.* will be referred to as 'revPBE-vdW' with the term 'vdW-DF' used when referring to the class of functionals. Results

using the OPLS-AA [193] potential for methane and the TIP4P-2005 [91] and TIP4P-ICE [114] potentials for water will also be presented. Details of these FFs are given in Appendix G, but key features of these potentials are that they are all-atomic, point charge and have Lennard-Jones sites located on the carbon and oxygen atoms. The TIP4P-2005 and TIP4P-ICE potentials are rigid, whereas the OPLS-AA potential is flexible. The two water potentials in combination with a number of different methane potentials have also been investigated, but as these yield similar results to OPLS-AA, they have been omitted from the main results for clarity and are included in Appendix G. Although this is clearly not an exhaustive list of possible *xc* functionals and FFs available, the test set nevertheless is adequate to highlight the main strengths and weaknesses of these types of methods in describing hydrogen-bond plus dispersion bound systems such as methane hydrate.

In the following sections, the results of the above mentioned *xc* correlation functionals and force fields to DMC in their prediction of the bulk properties of sI methane hydrate will be compared. Specifically, the cohesive energy of the hydrate crystal, the binding energy of the methane to the water framework and the dissociation energy of the hydrate crystal to ice I_h and methane vapour will be investigated. It will be seen that none of the methods give a particularly satisfactory description of bulk sI methane hydrate and that in instances of apparent agreement, this is due to a fortuitous cancellation of errors.

7.3 Computational Setup

DFT calculations were performed using `VASP 5.3.2` [194–196], a periodic plane-wave basis set code.[2] Calculations with the vdW-DFs have been carried out self-consistently using the scheme of Román-Pérez and Soler [197], as implemented in `VASP` by Klimeš *et al.* [59]. Projector-augmented-waves (PAW) potentials [198] have been used, with LDA-based potentials used for the LDA calculations and PBE-based potentials used for all other calculations. All results reported here used the 'standard' PAW potentials supplied with `VASP` and a plane-wave cutoff of 600 eV (these PAWs have been optimised for a plane-wave basis cutoff $\geq$ 400 eV).

[2]For the vdW^{TS} calculations, `VASP 5.3.3` was used.

A Γ-centred $2\times2\times2$ Monkhorst-Pack k-point mesh [199] per unit cell was used for calculations of bulk sI methane hydrate, whereas calculations concerning isolated molecules were performed at the Γ-point only, in a cubic simulation cell of volume $20 \times 20 \times 20\,\text{Å}^3$. The structures for bulk sI methane hydrate were taken from the work of Lenz and Ojmäe [178] and optimised using the conjugate gradient geometry optimiser until forces on all atoms were below 0.02 eV/Å. Wave functions were converged to within 1×10^{-8} eV. For calculations concerning ice I_h, the same proton-ordered 12 molecule ice I_h unit cell structures as those in Reference [176] have been used, with a Γ-centred $2 \times 2 \times 2$ Monkhorst-Pack k-point mesh per unit cell used. All other settings were identical to those used for the hydrate calculations.

All quantum Monte Carlo calculations were performed using version 2.12.1 of the Cambridge `CASINO` code [184]. DMC simulations for 178-atom simulation cells were performed using conventional Slater-Jastrow trial wave functions with a Jastrow factor containing electron-nucleus, electron-electron, and electron-nucleus electron terms [200], each of which depends on variational parameters determined by a combination of variance- and energy-minimization. The orbitals in the determinantal part of the trial wave function were generated from DFT calculations performed by the PWSCF component of the `QUANTUM ESPRESSO` package [201]; these Γ-point DFT calculations were done using the PBE xc functional and a 300 Ry (4082 eV) plane-wave cutoff. The same structures from Lenz and Ojamäe used for the `VASP` calculations were first optimised with these settings in the PWSCF component of the `QUANTUM ESPRESSO` package. As is standard practice, the plane-wave orbitals were re-expressed in B-splines [185] for the DMC simulations. Dirac-Fock pseudopotentials specifically developed for use in QMC were used [202, 203]. Although in principle pseudopotentials for hydrogen are not required, this would imply using e.g. Gaussian basis sets to construct the trial wavefunctions, the completeness of which is difficult to establish in a systematic way. Here it is preferred to use plane waves to achieve full, automatic and unbiased basis set convergence. The quality of the hydrogen pseudopotential is supported in References [187, 204, 205] where agreement is to within 3 meV/H_2O of CCSD(T) calculations and Reference [175], where agreement is to within 5 meV/H_2O of

experiment. Coulomb finite size effects were accounted for using the 'structure factor' method described in References [206] and [207] (Modified Periodic Coulomb (MPC) interaction defined in References [208] and [209] could have equally well been used, which when checked, gave essentially the same results).

FF calculations were performed using the `GROMACS 4.5.5` simulation package [65]. Long range electrostatics were treated with the particle-mesh Ewald method [210, 211] with a grid spacing of 1 Å used for the fast Fourier transform (fourth order interpolation was also used) and a real space cut-off of 9 Å. Lennard-Jones interactions were truncated after 9 Å with tail corrections applied. The calculations were also performed without the tail corrections and results from these have been included in Appendix G (any effect of the tail corrections does not alter the conclusions presented in this chapter). The `L-BFGS` algorithm [212, 213] was used to optimise the geometries, with the `SETTLES` algorithm [101] used to constrain the water geometry. All geometries were converged to within 1.05×10^{-6} eV/Å.

7.4 Results and Discussion

Gas hydrates come in three main crystal forms - structures I, II and H (sI, sII and sH, respectively). Methane hydrate is generally found in the sI form, although the sII and sH forms have been reported under very high pressure (above 250 MPa and *ca.* 1 GPa respectively) [214, 215]. In the sI hydrate, the water molecules form a hydrogen bonded network that gives rise to two types of cavities: a twelve-sided pentagonal dodecahedron (often denoted as 5^{12}); and a 14-sided tetrakaidecahedron (denoted as $5^{12}6^2$, owing to the fact that it consists of 12 pentagonal and 2 hexagonal faces). In stoichiometric sI hydrate, the methane molecules singly occupy each cavity. The cubic unit cell consists of two 5^{12} and six $5^{12}6^2$ cages and has the chemical formula $46H_2O \cdot 8CH_4$. The sI methane hydrate structure is shown in Figure 7.1(b) and a comprehensive overview of the sI, sII and sH hydrate structures can be found in Reference [3].

The cohesive energy per water molecule:

$$\Delta E_{\mathrm{coh}}^{\mathrm{sI}}(a) = \frac{E_{\mathrm{sI}}(a) - 46E_{\mathrm{H_2O}} - 8E_{\mathrm{CH_4}}}{46} \tag{7.16}$$

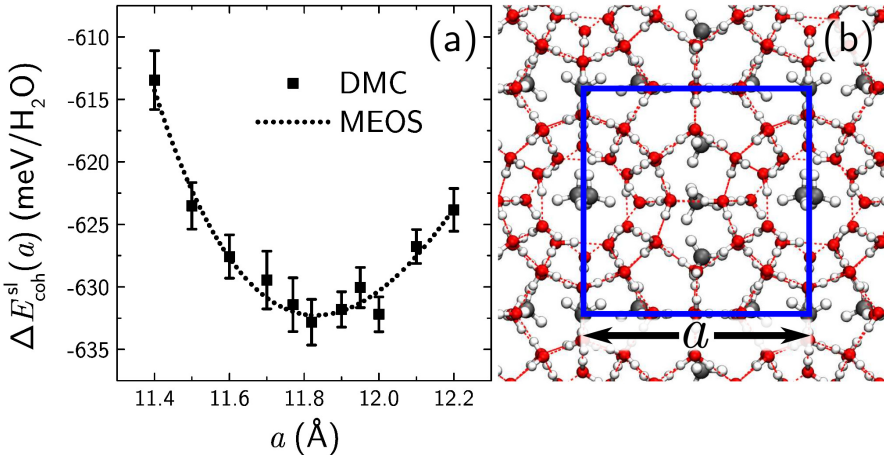

Figure 7.1: (a) Variation of the DMC cohesive energy of bulk sI methane hydrate with lattice constant. The cohesive energy is defined by Equation 7.16. The bars on each data point indicate a one standard deviation estimate of the error. From the fit to Murnaghan's equation of state (MEOS), the equilibrium lattice constant and cohesive energy are estimated to be 11.83 ± 0.02 Å and -632 ± 1 meV/H_2O, respectively. (b) Bulk sI methane hydrate crystal structure. The blue box bounds the unit cell. The atoms are coloured as: grey, carbon; red, oxygen; and white, hydrogen. The dashed red lines outline the hydrogen bonded water framework.

of the bulk sI methane hydrate unit cell has been computed for a variety of unit cell volumes, maintaining a cubic simulation cell. In Equation 7.16, $E_{\mathrm{sI}}(a)$ is the total energy of bulk sI methane hydrate with lattice constant a, whilst $E_{\mathrm{H_2O}}$ and $E_{\mathrm{CH_4}}$ are energies of the isolated water and methane molecules, respectively. This was done first using DMC. By fitting $\Delta E_{\mathrm{coh}}^{\mathrm{sI}}(a)$ to Murnaghan's equation of state [216, 217], the equilibrium lattice constant a_0 and cohesive energy $\Delta E_{\mathrm{coh}}^{\mathrm{sI}}(a_0)$ can be determined. These results are presented in Figure 7.1(a), where it can be seen that the equilibrium lattice constant is estimated to be 11.83 ± 0.02 Å and the cohesive energy is -632 ± 1 meV/H_2O. The DMC lattice constant compares well to the low temperature neutron scattering data of Davidson *et al.* [218] (11.77 ± 0.01 Å, CH_4/D_2O at 5.2 K) and Gutt *et al.* [219] (11.821 ± 0.001 Å, CD_4/D_2O at 2 K).

The variation of the cohesive energy with lattice constant for each of the DFT *xc* functionals and FFs discussed in Section 7.2 has also been computed. In these calculations, all atoms were allowed to relax independently (with the constraint of rigid water molecules for the FFs). The results of these calculations are presented in Figure 7.2 and Table 7.1 (the LDA results have been excluded from Figure 7.2 for clarity). Although all of the examined DFT *xc* function-

als overbind the hydrate crystal, there is considerable variety amongst the DFT results, with optPBE-, optB88-, optB86b-vdW, PBE-D2 and PBE-vdW$^{\mathrm{TS}}$ significantly overbinding the hydrate crystal, whilst PBE and revPBE-vdW yield cohesive energies in better agreement with DMC. Despite having the best agreement with DMC for the cohesive energy (within 1%), revPBE-vdW does, however, predict a lattice constant that is 1.9–2.3% too large. On the other hand, although it significantly overbinds the crystal, optPBE-vdW predicts a structure in decent agreement with DMC (0.6–0.9% too small), whereas optB88- and optB86b-vdW yield lattice constants that are too short by 2.3–2.6% and 2.5–2.9%, respectively. PBE-D2 and PBE-vdW$^{\mathrm{TS}}$ also strongly overbind the hydrate crystal and predict lattice constants that are too small by 3.0% or worse. What is perhaps surprising is that PBE, which fails to account for vdW interactions entirely, is yielding reasonable results not only for the structure (0.6–0.9% smaller than DMC), but also for the energetics. In fact, not only does PBE predict an equilibrium cohesive energy in reasonable agreement with the DMC result, it actually slightly overbinds the crystal by 2.1–2.4%. The force fields, OPLS-AA/TIP4P-2005 and OPLS-AA/TIP4P-ICE, overbind by 8.2-8.6% and 18.0-18.4% respectively, although their predicted structures are in decent agreement with the reference data, with their predicted lattice constants differing from DMC by less than 1.0%.

From the results for $\Delta E_{\mathrm{coh}}^{\mathrm{sI}}(a)$ presented in Figure 7.2 it would be tempting to conclude that PBE gives a satisfactory description of bulk sI methane hydrate. Given the well known problem that GGA functionals do not account for dispersion interactions, however, the fact that PBE slightly overbinds the hydrate seems almost paradoxical. Furthermore, the overly repulsive nature of revPBE exchange at short separations has been shown to lead to lattice constants that are too long and cohesive energies that are too weak in hydrogen bonded systems such as ice [176]. Indeed, a lattice constant with revPBE-vdW that is 1.9–2.3% too large is obtained, but why then, is the cohesive energy for sI methane hydrate slightly too strong with this functional? To better understand these results, the total cohesive energy has been decomposed into contributions arising from the methane binding

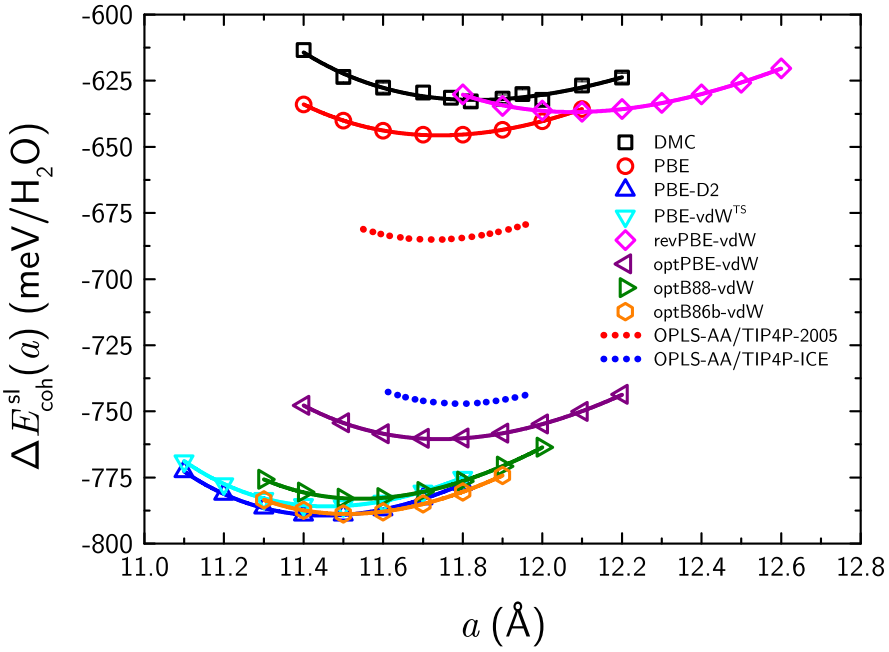

Figure 7.2: Variation of the cohesive energy of bulk sI methane hydrate with lattice constant. The cohesive energy is defined by Equation 7.16. Symbols represent calculated values using DFT (empty squares show DMC data), whereas the solid lines show a fit to Murnaghan's equation of state. For the DMC data, the error bars are smaller than the size of the symbols. Results using the OPLS-AA force field for methane in combination with the TIP4P-2005 and TIP4P-ICE water potentials are also shown (fit to Murnaghan's equation of state only). The results for LDA, which has a cohesive energy of $-1178\,\mathrm{meV/H_2O}$ and equilibrium lattice constant of 10.933 Å, have been omitted for clarity.

to the empty hydrate:

$$\Delta E_{\mathrm{CH_4}} = \frac{E_{\mathrm{sI}}(a_0) - E_{\mathrm{empty}}(a_0) - 8E_{\mathrm{CH_4}}}{8} \tag{7.17}$$

and the cohesive energy of the empty hydrate:

$$\Delta E_{\mathrm{coh}}^{\mathrm{empty}} = \frac{E_{\mathrm{empty}}(a_0) - 46E_{\mathrm{H_2O}}}{46}, \tag{7.18}$$

where $E_{\mathrm{empty}}(a_0)$ is the energy of the hydrate unit cell with no methane present, calculated without further relaxation of the water molecules (i.e. the water molecules are 'frozen' in the position they assume in the bulk hydrate). For DMC, both $\Delta E_{\mathrm{CH_4}}$ and $\Delta E_{\mathrm{coh}}^{\mathrm{empty}}$ have been calculated at the experimental lattice constant [218] $a = 11.77\,$Å. These results are presented in Figure 7.3 and Table 7.1, with DMC providing reference values of $\Delta E_{\mathrm{CH_4}} = -241 \pm 15\,\mathrm{meV/CH_4}$

and $\Delta E_{\mathrm{coh}}^{\mathrm{empty}} = -590 \pm 2\,\mathrm{meV/H_2O}$. The origin of PBE's seemingly good description of bulk sI hydrate now becomes apparent: the lack of vdW interactions means there is no binding between the methane and the water (in fact $\Delta E_{\mathrm{CH_4}}$ is slightly positive), but this is compensated for by an overbinding of the hydrogen bonded water framework. Although the overbinding of the water framework is small on a per molecule basis, water and methane exist in a ratio of 23:4 in the stoichiometric hydrate, meaning that small errors in describing the water-water interactions are much amplified compared to the apparently larger errors in the methane binding energy. From Figure 7.3 it can also be seen that LDA's severe overbinding occurs principally from its description of the water framework ($\Delta E_{\mathrm{coh}}^{\mathrm{empty,LDA}} - \Delta E_{\mathrm{coh}}^{\mathrm{empty,DMC}} = -531\,\mathrm{meV/H_2O}$), although it is worth noting that it also overbinds the methane to the water framework by $85\,\mathrm{meV/CH_4}$. The ability of LDA to bind van der Waals systems (such as CH_4 in a H_2O cage) has been observed before [220, 221]; this is known to be fortuitous because, by its nature, LDA relies on a local description of exchange and correlation and does not account for non-local interactions. Turning now to the dispersion-corrected functionals it is clear that, with the exception of PBE-D2, they all over-correct the neglect of vdW interactions by the GGA functional, yielding methane binding energies that are too strong by 138–262 $\mathrm{meV/CH_4}$. It is also clear that the better agreement of the cohesive energy obtained with revPBE-vdW compared to the other dispersion-corrected functionals is due to an underbinding of the water framework (consistent with results obtained for bulk ice I_h [176]) that offsets a strong overbinding of the methane. In the case of the other dispersion-corrected functionals, as well as predicting methane binding energies that are too exothermic, they also overbind the water framework by 83–154 $\mathrm{meV/H_2O}$. The source of overbinding for the FFs occurs almost exclusively in the water framework, with both FFs presented here yielding good agreement for $\Delta E_{\mathrm{CH_4}}$.

Due to the high water content of sI methane hydrate, it is convenient to compare the performance of the *xc* functionals and FFs for the hydrate to ice I_h. Ice I_h is chosen rather than any of the other phases of ice due to its close structural similarity to sI hydrate at the molecular level: the average hydrogen bond length in the hydrate is only 1% longer on average than in ice I_h and the

Table 7.1: Computed cohesive energies $\Delta E_{\mathrm{coh}}^{\mathrm{sI}}(a_0)$, equilibrium lattice constants (a_0), methane binding energies to the empty hydrate ($\Delta E_{\mathrm{CH_4}}$), empty hydrate cohesive energies ($\Delta E_{\mathrm{coh}}^{\mathrm{empty}}$), ice cohesive energies ($\Delta E_{\mathrm{coh}}^{\mathrm{ice}}$), and methane hydrate dissociation energies to ice $\mathrm{I_h}$ and gas ($\Delta E_{\mathrm{diss}}^{\mathrm{sI}\rightarrow\mathrm{ice}}$). The DMC value for $\Delta E_{\mathrm{coh}}^{\mathrm{ice}}$ is taken from Reference [175]. The unit of $\Delta E_{\mathrm{coh}}^{\mathrm{sI}}(a_0)$, $\Delta E_{\mathrm{coh}}^{\mathrm{empty}}$ and $\Delta E_{\mathrm{coh}}^{\mathrm{ice}}$ are meV/$\mathrm{H_2O}$, whilst $\Delta E_{\mathrm{CH_4}}$ and $\Delta E_{\mathrm{diss}}^{\mathrm{sI}\rightarrow\mathrm{ice}}$ are given in meV/$\mathrm{CH_4}$. The equilibrium lattice constant a_0 is given in Ångstrom.

Method	$\Delta E_{\mathrm{coh}}^{\mathrm{sI}}(a_0)$	a_0	$\Delta E_{\mathrm{CH_4}}$	$\Delta E_{\mathrm{coh}}^{\mathrm{empty}}$	$\Delta E_{\mathrm{coh}}^{\mathrm{ice}}$	$\Delta E_{\mathrm{diss}}^{\mathrm{sI}\rightarrow\mathrm{ice}}$
LDA	−1178	10.933	−326	−1121	−1136	+240
PBE	−646	11.740	+15	−648	−657	−67
PBE-D2	−789	11.453	−262	−744	−758	+179
PBE-vdW$^{\mathrm{TS}}$	−786	11.461	−379	−720	−737	+280
revPBE-vdW	−637	12.077	−423	−563	−583	+308
optPBE-vdW	−760	11.743	−503	−673	−696	+369
optB88-vdW	−783	11.542	−468	−702	−725	+335
optB86b-vdW	−789	11.509	−458	−709	−733	+321
OPLS-AA/TIP4P-2005	−685	11.726	−248	−642	−653	+184
OPLS-AA/TIP4P-ICE	−747	11.792	−261	−702	−714	+190
DMC	-632 ± 1	11.83 ± 0.02	-241 ± 15	-590 ± 2	-605 ± 5	$+155 \pm 34$

hydrate O–O–O angles differ from the tetrahedral angles of ice $\mathrm{I_h}$ by only 3.7° [3]. In the same manner that $\Delta E_{\mathrm{coh}}^{\mathrm{sI}}(a_0)$ was calculated for sI methane hydrate, $\Delta E_{\mathrm{coh}}^{\mathrm{ice}}$ has also been computed for the test set of *xc* functionals and FFs by fitting the cohesive energy of the bulk ice $\mathrm{I_h}$ crystal to Murnaghan's equation of state (a more comprehensive overview of the ice results is given in Appendix G). For DMC, the value of $\Delta E_{\mathrm{coh}}^{\mathrm{ice}}$ was taken from Reference [175]. From these calculations, the equilibrium volume of the ice $\mathrm{I_h}$ crystal was also obtained. In Figure 7.4 the difference in computed volume using DFT/FF from that using DMC is shown for sI methane hydrate, plotted against the same quantity for ice $\mathrm{I_h}$. There is a strong correlation between the errors in computed volumes for the hydrate and ice $\mathrm{I_h}$, suggesting that the primary factor in obtaining reasonable lattice volumes for the hydrate is an accurate description of the hydrogen bonded water framework. Figure 7.4 also shows the differences in DFT/FF values for $\Delta E_{\mathrm{coh}}^{\mathrm{sI}}$ and $\Delta E_{\mathrm{coh}}^{\mathrm{empty}}$ from DMC, again plotted against the DFT/FF-DMC difference for the ice $\mathrm{I_h}$ cohesive energy. As for the volumes, a very strong positive correlation is seen between the errors in the hydrate cohesive energies and those for ice $\mathrm{I_h}$. In fact, there is a near perfect correlation for the error $\Delta E_{\mathrm{coh}}^{\mathrm{empty}}$ and the error in $\Delta E_{\mathrm{coh}}^{\mathrm{ice}}$, the significance of which will become apparent when the dissociation behaviour

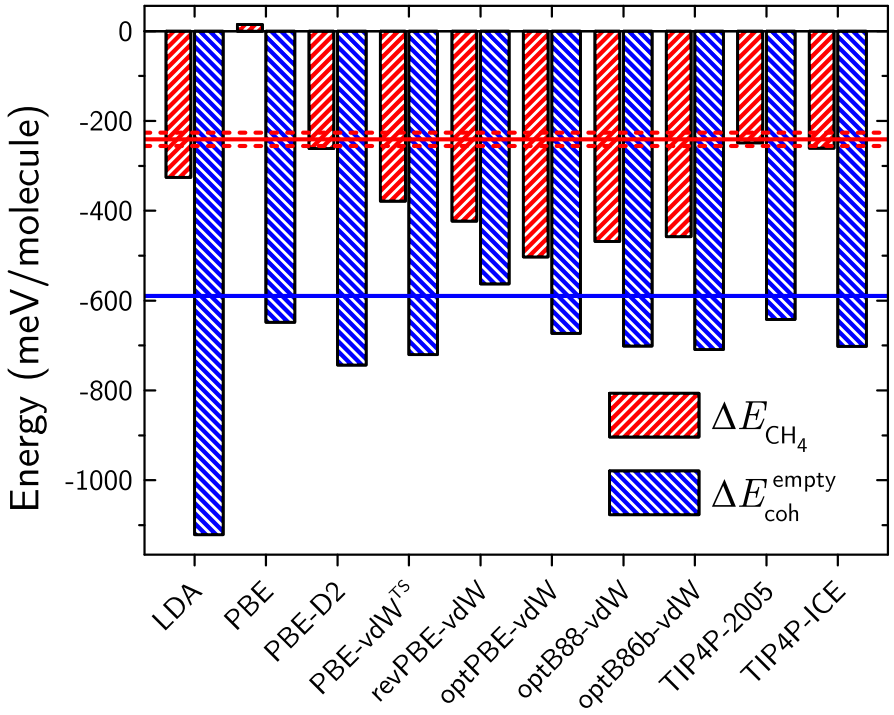

Figure 7.3: Binding energy of methane to the empty hydrate $\Delta E_{\mathrm{CH_4}}$ and formation energy of the empty hydrate $\Delta E_{\mathrm{coh}}^{\mathrm{empty}}$. The horizontal red and blue lines show the DMC values for $\Delta E_{\mathrm{CH_4}}$ and $\Delta E_{\mathrm{coh}}^{\mathrm{empty}}$ respectively (dashed lines show the associated statistical uncertainty - not visible for $\Delta E_{\mathrm{coh}}^{\mathrm{empty}}$). Apart from PBE-D2, all dispersion-corrected density functionals severely overbind methane to the empty hydrate. Similarly, all density functionals overbind the water framework, with the exception of revPBE-vdW. PBE, which does not account for vdW interactions, fails to predict methane binding to the empty hydrate structure. The force fields yield good values for $\Delta E_{\mathrm{CH_4}}$, but like the DFT methods, they overbind the water framework.

of the hydrate to ice $\mathrm{I_h}$ and methane gas is considered.

In comparing the cohesive energies of the DFT xc functionals and point charge FFs to DMC, the vapour phase of both methane and water is being taken as the reference state. More important to the phase equilibria of gas hydrates, however, is the relative energy of the hydrate with respect to methane gas and another condensed phase of water, either liquid or ice [3]. Whilst the cancellation of errors in $\Delta E_{\mathrm{CH_4}}$ and $\Delta E_{\mathrm{coh}}^{\mathrm{empty}}$ means that PBE has a good overall agreement with the DMC cohesive energy, it is straightforward to demonstrate that the error in $\Delta E_{\mathrm{CH_4}}$ arising from the neglect of vdW interactions can lead to severe consequences regarding the thermodynamic stability of sI methane hydrate. Consider the process of sI methane hydrate dissociating to ice $\mathrm{I_h}$ and methane gas:

$$5.75\mathrm{H_2O} \cdot \mathrm{CH_{4(sI)}} \xrightarrow{\Delta E_{\mathrm{diss}}^{\mathrm{sI}\rightarrow\mathrm{ice}}} 5.75\mathrm{H_2O_{(ice)}} + \mathrm{CH_{4(gas)}}. \tag{7.19}$$

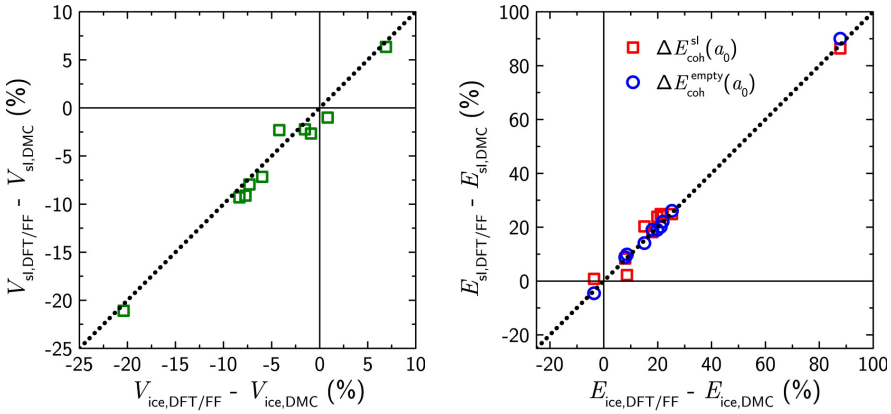

Figure 7.4: Comparison of *xc* functional and FF performance for ice I_h and sI methane hydrate. The left panel shows the percentage difference from the DMC sI hydrate volume against the percentage difference from the DMC ice I_h volume, for the various DFT *xc* functionals and force fields. The right panel shows the percentage difference from the DMC sI hydrate cohesive energies (ΔE_{coh}^{sI} and ΔE_{coh}^{empty}) against the percentage difference from the DMC cohesive energy for ice I_h, again for all the *xc* functionals and FFs investigated.

The associated energy cost $\Delta E_{diss}^{sI \to ice}$ can be computed as (see Appendix G):

$$\Delta E_{diss}^{sI \to ice} = 5.75 \Delta E_{coh}^{ice} - 5.75 \Delta E_{coh}^{empty} - \Delta E_{CH_4}. \tag{7.20}$$

The results of these calculations are presented in Table 7.1. It is clear that the results for PBE are disastrous: sI methane hydrate is unstable with respect to dissociation to ice I_h and methane gas by 67 meV/CH_4 (i.e. it is 67 meV/CH_4 exothermic). In contrast, DMC predicts dissociation to be an endothermic process, costing 155 ± 34 meV/CH_4. It is noted here that the experimental enthalpy of dissociation [222] at standard temperature and pressure is 188 ± 3 meV/CH_4 suggesting that the DMC value is reasonable.[3] All of the dispersion-corrected functionals improve on the GGA functional in this respect, predicting that the hydrate is stable with respect to ice and methane gas. PBE-D2 gives the best

[3] As the experimental number is a standard enthalpy of dissociation we should not expect quantitative agreement with the DMC dissociation energy, which is a total energy difference. Aside from the temperature/pressure effects present in experiment, there is also the issue of non-stoichiometry (the experimental data of Handa [222] was obtained for a methane occupancy of water cages of *ca.* 96%), which means that configurational entropy is likely to be important for the experimental dissociation enthalpy. This comparison is made simply to show that the number obtained with DMC is reasonable. In fact, analysis of numerous experimental data sets using the Clapeyron equation yields an enthalpy of dissociation of 157 ± 6 meV/CH_4 at 150 K and 0.0564 bar [223]. This is arguably a better comparison to the zero temperature/pressure DMC calculations and indeed improves agreement, but one should nevertheless exercise caution when comparing a calculated dissociation energy to an experimental enthalpy.

agreement with DMC, followed by LDA and PBE-vdW$^{\mathrm{TS}}$, although it should be kept in mind that these calculations have been performed at the equilibrium volume of the *xc* functional used. The vdW-DFs over-stabilise the hydrate by a factor of approximately two. Unsurprisingly, the trends in $\Delta E_{\mathrm{coh}}^{\mathrm{ice}}$ closely follow those of $\Delta E_{\mathrm{CH_4}}$, with the errors in describing the hydrogen bonded water network more-or-less cancelling between $\Delta E_{\mathrm{coh}}^{\mathrm{empty}}$ and $\Delta E_{\mathrm{coh}}^{\mathrm{ice}}$ (as shown in Figure 7.4). As such, the FFs also give good agreement with the DMC result. The fact that the point charge FFs predict $\Delta E_{\mathrm{coh}}^{\mathrm{sI}}$ to be too exothermic can be attributed to the enhanced dipole moment of the isolated water molecules in these types of potentials [63, 224], which has been shown to lead to too high vaporisation enthalpies of ice $\mathrm{I_h}$ for the TIP4P-2005 and TIP4P-ICE potentials [225]. Indeed, Vega and co-workers [225] have found that it is impossible to simultaneously fit the melting temperature of ice $\mathrm{I_h}$ and the enthalpy of vaporisation for such models. It is therefore probably expecting too much of the rigid point charge FFs to give reasonable results for both $\Delta E_{\mathrm{coh}}^{\mathrm{sI}}$ and $\Delta E_{\mathrm{diss}}^{\mathrm{sI}\rightarrow\mathrm{ice}}$ whilst also maintaining favourable densities and coexistence/melting temperatures for the hydrate and ice $\mathrm{I_h}$ [226]. Use of an explicitly polarizable water potential may go some way to improving this situation [227].

7.5 Conclusions

High-quality DMC reference data for bulk sI methane hydrate has been presented and the performance of several commonly used *xc* functionals and point charge force fields evaluated. It has been found that none of the DFT methods tested give particularly satisfactory results. Van der Waals forces have been found to be crucial to the stability of methane hydrate with respect to dissociation to ice $\mathrm{I_h}$ and methane gas, although the vdW-DF flavour of *xc* functionals over-stabilise the hydrate by approximately a factor of two. This effect is less severe with the PBE-D2 and PBE-vdW$^{\mathrm{TS}}$ functionals, although their equilibrium volumes are too small compared to DMC. PBE, which neglects dispersion interactions, incorrectly predicts that methane hydrate is unstable with respect to dissociation to ice and methane gas. By overbinding the hydrogen bonded water framework, PBE's poor description of the water-methane interaction is compensated, giving a good overall

agreement with the DMC cohesive energy of the bulk hydrate. This last point highlights the difficulty that DFT *xc* functionals face in describing mixed phase systems such as gas hydrates; in order to obtain a good overall description, it is necessary to be able to accurately describe both the hydrogen bonded water framework and the dispersion bound methane. It has also been seen that point-charge, all-atom force fields tend to overbind the hydrate lattice, although their agreement with DMC for the dissociation energy to ice and vapour, and for the structure for the bulk crystal, is good. From knowledge of the literature [225] on the performance of simple point charge FFs for ice, it is unlikely that such FFs will be able to simultaneously describe both the cohesive energy of the hydrate crystal and the energetics of dissociation to other condensed phase water systems.

Earlier in this chapter, it was remarked that the 23:4 ratio of water to methane amplified the apparently small errors in the water-water interactions compared to the water-methane interactions. It was also seen that the high water content means that the errors in describing the hydrate are strongly correlated to the errors in describing ice I_h. However, such a high water-methane ratio also means that there is the possibility for significant many body interactions between the methane and water (e.g. a single isolated 5^{12} cage has 190 water-methane-water triplets). Indeed, a separate independent study investigating the binding energy of methane to a gas phase 5^{12} cage through a many-body expansion of the total energy has found significant contributions to the DFT error beyond those in the two-body interactions (symmetry adapted perturbation theory calculations also showed that the DFT methods have insufficiencies other than those associated with the neglect of long-range dispersion interactions) [228].

Although the effects of including exact exchange have not been considered here, it is unlikely that this will significantly improve the DFT description of sI methane hydrate. For example, using the PBE0 [229] results for ice I_h from Reference [176], and for a methane molecule binding to a gas phase 5^{12} water cage from Reference [228], it can be predicted that this hybrid *xc* functional will give a reasonable prediction of the hydrate structure (similar to PBE) but will still incorrectly destabilise the hydrate with respect to methane gas and ice I_h. Including dispersion corrections to this functional, such as PBE0-D2 or PBE0-vdW^{TS}, can

therefore be expected to also give similar results to PBE-D2 and PBE-vdW$^{\mathrm{TS}}$. It has recently been shown [230] that accurate DMC reference data in combination with Gaussian approximation potentials [231] can be used to systematically correct the 'beyond two-body' errors associated with GGA functionals for water nano-droplets and bulk liquid water. Such an approach is also likely to be more successful in improving the performance of DFT *xc* functionals for gas hydrates compared to the pairwise additive dispersion corrections examined here.

Chapter 8

Summary and Outlook

In this thesis, a variety of computational techniques have been used to investigate heterogeneous ice nucleation. Grand canonical Monte Carlo in the presence of simplified model surfaces was used to test assumptions of a long standing theory of the role of a surface's lattice mismatch in ice nucleation. Contrary to conventional wisdom, it was found that even at hexagonal surfaces, water readily adopts structures that deviate from the hexagonal motifs found in ice. This was due to the competition of adsorbing strongly to the surface whilst also maintaining favourable water–water interactions.

Heterogeneous nucleation at a more realistic surface model of a clay mineral was then investigated, using brute force molecular dynamics simulations. It was found that the nature of the hydrogen bonding with the surface played an important role in how ice formed at the surface: it was always found that ice nucleated with its prism face bound to the clay mineral, as this face provided a means to form hydrogen bonds not only to the surface, but also to other water molecules. This result was somewhat surprising, as the clay mineral in question was thought to aid ice formation by acting as a template for the basal face of ice. It was also speculated that the ability of a surface to induce ice formation along one particular face could affect the macroscopic morphology of ice crystals.

Rates of heterogeneous ice nucleation in the presence of different types of model nanoparticles, using a coarse grained potential, were then computed. The results obtained represent the single biggest advance in this thesis in terms of

understanding how the properties of a surface affect ice nucleation. It was found that just by changing its hydrophilicity, a nanoparticle could either inhibit or promote ice nucleation. This observation was explained by the molecular insight extracted from the simulations: it was found that the decrease in the rate with stronger adsorption to the surface was due to the presence of excess sites – sites at the surface not occupied when ice forms. Modification of the surface to block these excess sites led to an increased rate of ice formation. Another finding of these simulations was that heterogeneous nucleation can proceed by distinctly different mechanisms, depending upon the topography of the surface exposed to water. This directly challenges the notion of the often used 'requirements' of an ice nucleating agent: a property that aides ice nucleation on one surface could be detrimental for another.

A look at one of the more fundamental aspects of ice nucleation was then investigated using transition path sampling techniques. In particular, the role of dynamic heterogeneity in homogeneous ice nucleation was studied. It was found that ice formation occurs in pre-existing immobile domains of water molecules. At higher temperatures, this was accompanied by a decrease in the number of mobile particles prior to ice formation. The TPS simulations also found no evidence to support the claim that a newly proposed phase 'ice 0' plays an important role in homogeneous ice nucleation.

Finally, the accuracy of commonly used DFT exchange-correlation functionals in describing bulk sI methane hydrate was examined. It was found that none of the functionals examined adequately described the hydrate and that the errors correlated well with the errors in describing ice I_h. The results highlighted the difficulty for DFT in describing both the dispersion bound methane and hydrogen bonded water framework.

Even though advances have been made in this thesis in furthering our understanding of heterogeneous ice nucleation, challenges certainly remain. In Chapter 2, it was remarked that Equation 2.20:

$$Z = V^{-N} \int \mathrm{d}\mathbf{r}^N \exp\left(-\beta U(\mathbf{r}^N)\right)$$

encompasses essentially all of computational chemistry. To fully understand heterogeneous ice nucleation will require both challenges captured in this relationship – the sampling of configuration space and the accuracy of the intermolecular potential – to be addressed simultaneously. Currently, the coarse grained potential offers the best way to study ice nucleation, whether or not enhanced sampling techniques are used. This situation needs to change, as such a simple model can only be pushed so far. The intricacies of the hydrogen bonding between the surface and the water, for example, cannot be captured. To model systems of relevance to atmospheric chemistry or industry, and to be able to make quantitative predictions, will require all atom models. The development of enhanced sampling techniques to deal with the slow dynamics of such potentials is therefore needed. Given the difficulty in identifying reaction coordinates for ice nucleation, it is a personal opinion that transition path sampling and forward flux sampling techniques offer the best way forward. Even when nucleation studies with all atom potentials and reasonable system sizes become routine, a major challenge will be to design realistic surface models, as many surfaces reconstruct, adopting structures somewhat different to those obtained simply from bulk truncation. Having an adequate description of the intermolecular potential between water and surface defects also needs to be considered, which will require high-level *ab initio* calculations to at least provide accurate benchmark data.

The above challenges should not instill pessimism in the reader. Advances in computer architecture and algorithms have meant that molecular simulations that were challenging ten years ago are now routine, and there is no reason why this trend should not continue. The bulk of the machinery needed to tackle heterogeneous ice nucleation with molecular simulation is available (e.g. enhanced sampling techniques, systematic ways to improve intermolecular potentials and structure searching methods) but requires some fine tuning. With some effort, simulations of heterogeneous ice nucleation on realistic model surfaces should become routine in the next few years. It is hoped that the work presented in this thesis has gone someway to helping this process already.

Appendix A

From Quantum to Classical

In Chapter 2, we derived many of the Statistical Mechanical formulae in the context of a quantum mechanical system. A classical approximation for the canonical partition function was then invoked in Section 2.1.4:

$$Q(\beta) = \mathrm{Tr}\left[\mathrm{e}^{-\beta\hat{H}}\right] \rightarrow \frac{1}{h^{dN}N!}\int \mathrm{d}\mathbf{r}^N \mathrm{d}\mathbf{p}^N \exp\left(-\beta H(\mathbf{r}^N, \mathbf{p}^N)\right). \tag{2.17}$$

In this appendix, how we arrive at this classical approximation is outlined, following closely the derivation given in Reference [25].

We have already seen in Equation 2.15 that thermal averages of the quantum mechanical system are computed as:

$$\langle X\rangle = \frac{\sum_i \exp\left(-\beta E_{\mathrm{sys}}^{(i)}\right)\langle i|\,\hat{X}\,|i\rangle}{\sum_i \exp\left(-\beta E_{\mathrm{sys}}^{(i)}\right)}, \tag{2.15}$$

which can be recast as:

$$\begin{aligned}\langle X\rangle &= \frac{\sum_i \langle i|\exp\left(-\beta\hat{H}\right)\hat{X}\,|i\rangle}{\sum_i \langle i|\exp\left(-\beta\hat{H}\right)|i\rangle} \\ &= \frac{\mathrm{Tr}\left[\exp\left(-\beta\hat{H}\right)\hat{X}\right]}{\mathrm{Tr}\left[\exp\left(-\beta\hat{H}\right)\right]}.\end{aligned} \tag{A.1}$$

As the trace of a matrix is independent of the choice of basis, we can choose

any complete set that is convenient, such as the eigenstates of the position or momentum operator ($\hat{H} = \hat{K} + \hat{U}$, where $\hat{K}$ and $\hat{U}$ are the kinetic and potential energy operators, respectively). As the kinetic energy is the sum of the momenta squared of all the particles, the momentum eigenstates are also eigenfunctions of $\hat{K}$. As $\hat{U}$ is a function only of the particle positions, its matrix elements are most conveniently expressed in the basis set of position eigenstates. The total Hamiltonian $\hat{H} = \hat{K} + \hat{U}$, however, is not diagonal in either of these basis sets. It is also the case that:

$$\exp\left(-\beta\hat{H}\right) = \exp\left(-\beta(\hat{K} + \hat{U} + [\hat{K}, \hat{U}])\right), \tag{A.2}$$

where $[\hat{K}, \hat{U}]$ is the commutator of the kinetic and potential energy operators:

$$\begin{aligned}[\hat{K}, \hat{U}] &= \hat{K}\hat{U} - \hat{U}\hat{K} \\ &= \frac{i\hbar}{m}\left[(\nabla U(\mathbf{r}^N))\hat{p} - 2i\hbar\nabla^2 U(\mathbf{r}^N)\right],\end{aligned} \tag{A.3}$$

where $\hat{p} = -i\hbar\nabla$ is the momentum operator. As $[\hat{K}, \hat{U}]$ is on the order of $\hbar$, then it can be ignored in the classical limit, allowing us to write:

$$\mathrm{Tr}\left[\exp\left(-\beta\hat{H}\right)\right] \approx \mathrm{Tr}\left[\exp\left(-\beta\hat{K}\right)\exp\left(-\beta\hat{U}\right)\right]. \tag{A.4}$$

In the following, we will consider a one dimensional case for simplicity. Let us calculate $\mathrm{Tr}\left[\exp\left(-\beta\hat{H}\right)\right]$ in the basis set $\{|x\rangle\}$ of eigenstates of the position operator $\hat{x}$:

$$\begin{aligned}\mathrm{Tr}\left[\exp\left(-\beta\hat{H}\right)\right] &= \int \mathrm{d}x \,\langle x|\exp\left(-\beta\hat{K}\right)\exp\left(-\beta\hat{U}\right)|x\rangle \\ &= \int \mathrm{d}x \,\langle x|\exp\left(-\beta\hat{K}\right)|x\rangle \exp\left(-\beta U(x)\right).\end{aligned} \tag{A.5}$$

Now use the completeness relation:

$$\int \mathrm{d}a \,|a\rangle\langle a| = 1 \tag{A.6}$$

for the normalised basis set of plane waves:

$$\langle x|p\rangle = \frac{1}{\sqrt{2\pi\hbar}} \mathrm{e}^{ipx/\hbar}. \tag{A.7}$$

We can insert this into Equation A.5:

$$\begin{aligned} \mathrm{Tr}\left[\exp\left(-\beta\hat{H}\right)\right] &= \int \mathrm{d}x\,\mathrm{d}p\,\langle x|\exp\left(-\beta\hat{K}\right)|p\rangle\,\langle p|x\rangle\exp\left(-\beta U(x)\right) \\ &= \int \mathrm{d}x\,\mathrm{d}p\,|\,\langle x|p\rangle\,|^2\exp\left(-\beta K(p)\right)\exp\left(-\beta U(x)\right) \\ &= \int \frac{\mathrm{d}x\,\mathrm{d}p}{h}\exp\left(-\beta H(x,p)\right) \end{aligned} \tag{A.8}$$

Equation A.8 is the one-dimensional equivalent of Equation 2.17.

Appendix B

Supplementary Material for Chapter 3

This appendix provides supplementary material for Chapter 3. This includes details of the two model surfaces used, along with additional results not presented in the main chapter.

B.1 Honeycomb Surface

The surface was represented by a hexagonal close packed array of Lennard-Jones sites, separated by a lattice constant $a_{0,\mathrm{S}}$, consisting of four layers (see Figure B.1). The parameters used in this LJ potential are $\sigma_{\mathrm{OS}} = 3.34596\,\text{Å}$ and $\epsilon_{\mathrm{OS}} = 6.990697\,\mathrm{kJ/mol}$ and were chosen to give monomer adsorption energies that were comparable to the TIP4P dimer binding energy ($-26.09\,\mathrm{kJ/mol}$) [57]. Keeping the LJ parameters fixed for different lattice spacings did lead to only a slight change in monomer adsorption energies (see Table B.1). To test whether or not this change in binding energy significantly affects the results, a 'trial-and-error' approach was applied to varying the Lennard-Jones well depth until the monomer binding energy of the $\delta = 0.07$ surface was approximately that of the $\delta = 0.00$ surface. A GCMC simulation was then performed on this modified $\delta = 0.07$ surface at $\mu = -49.8\,\mathrm{kJ/mol}$ and the resulting structure analysed, averaging over a 15 K canonical MC simulation. The radial distribution functions of the original and modified surfaces are given in Figure B.2, where it can be seen

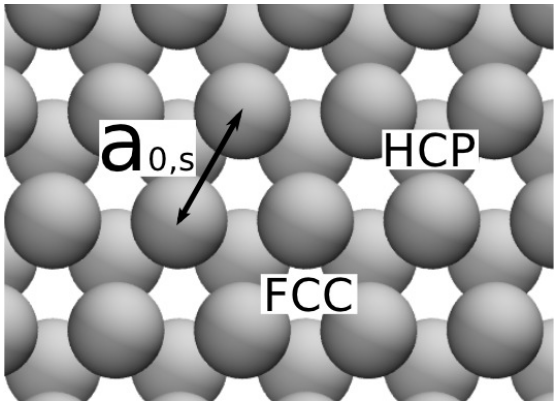

Figure B.1: Schematic of the model honeycomb surface. The surface lattice constant ($a_{0,\mathrm{S}}$) is the nearest neighbour distance between coplanar sites. The potential minima for monomer adsorption are found at the HCP and FCC sites labelled in the diagram.

that the structures are, to all intents and purposes, identical. This gives us reason to believe that any change in structures seen at the other surfaces is due to the change in lattice constant and not due to any change in binding energy to the surface.

δ	$a_{0,\mathrm{S}}$ (Å)	E_{bind} (kJ/mol)
-0.14	3.818	-32.93 ± 0.05
-0.07	4.307	-30.42 ± 0.05
0.00	4.440	-28.64 ± 0.05
0.04	4.618	-28.44 ± 0.07
0.07	4.751	-27.36 ± 0.05

Table B.1: Binding energies for a water monomer on the different honeycomb surfaces.

B.2 HCP Surface

The external potential $V_{\mathrm{ext}}(x_{\mathrm{ox}}, y_{\mathrm{ox}}, z_{\mathrm{ox}})$ described qualitatively in the main text can be defined more formally as:

$$V_{\mathrm{ext}}(x_{\mathrm{ox}}, y_{\mathrm{ox}}, z_{\mathrm{ox}}) = F(x_{\mathrm{ox}}, y_{\mathrm{ox}}) V_{\mathrm{Morse}}(z_{\mathrm{ox}}), \tag{B.1}$$

where $x_{\mathrm{ox}}, y_{\mathrm{ox}}, z_{\mathrm{ox}}$ are the coordinates of the water oxygen atom, $V_{\mathrm{Morse}}(z_{\mathrm{ox}})$ is a Morse potential dependent only on the normal distance of the water molecule from the surface and $F(x_{\mathrm{ox}}, y_{\mathrm{ox}})$ is a modulating function, dependent on the lateral

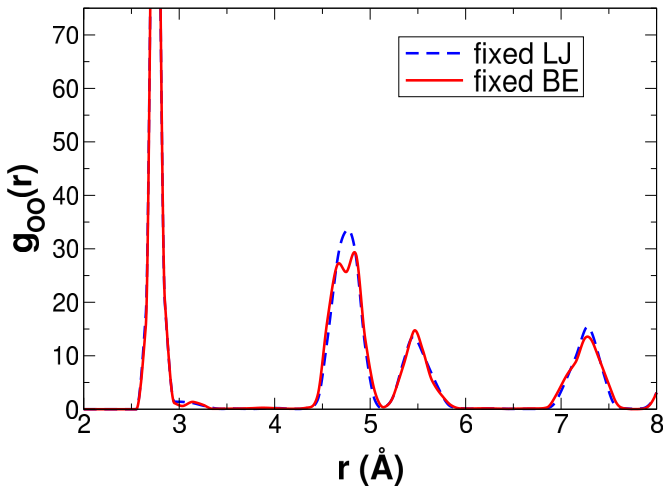

Figure B.2: Radial distribution function comparison of the $\delta = 0.07$ surface with fixed Lennard-Jones parameters (dashed blue line) and with the Lennard-Jones parameters increased (solid red line) to give a binding energy approximately that of the $\delta = 0.00$ surface (28.65 and 28.64 kJ/mol respectively).

displacement of the water molecule. The explicit forms of these functions are:

$$V_{\text{Morse}}(z_{\text{ox}}) = D_{z0}\{1 - \exp\left[-k_z(z_{\text{ox}} - z_0)\right]\}^2 - D_{z0}; \tag{B.2}$$

$$F(x_{\text{ox}}\,,\, y_{\text{ox}}) = \delta E_{\text{frac}}\left[\left(\frac{1-\delta E_{\text{frac}}}{\delta E_{\text{frac}}} - f(x_{\text{ox}}\,,\, y_{\text{ox}})\right)\right]; \tag{B.3}$$

and

$$\begin{aligned} f(x_{\text{ox}}\,,\, y_{\text{ox}}) = &\left(1 - \exp\left[-k_{xy}\left(x_{\text{ox}}^2 + y_{\text{ox}}^2\right)^{\frac{1}{2}}\right]\right)^2 - 1 \\ &+ \left(1 - \exp\left[-k_{xy}\left((x_{\text{ox}} - x_1)^2 + (y_{\text{ox}} - y_1)^2\right)^{\frac{1}{2}}\right]\right)^2 - 1 \\ &+ \left(1 - \exp\left[-k_{xy}\left((x_{\text{ox}} - x_1)^2 + (y_{\text{ox}} + y_1)^2\right)^{\frac{1}{2}}\right]\right)^2 - 1 \\ &+ \left(1 - \exp\left[-k_{xy}\left((x_{\text{ox}} + x_1)^2 + (y_{\text{ox}} - y_1)^2\right)^{\frac{1}{2}}\right]\right)^2 - 1 \\ &+ \left(1 - \exp\left[-k_{xy}\left((x_{\text{ox}} + x_1)^2 + (y_{\text{ox}} + y_1)^2\right)^{\frac{1}{2}}\right]\right)^2 - 1. \end{aligned} \tag{B.4}$$

where D_{z0} is the well depth of the potential perpendicular to the surface, k_z controls the width of the potential and z_0 is the position of the potential energy minimum above the surface. The function $f(x_{\text{ox}}\,,\, y_{\text{ox}})$ is composed of five symmetrised Morse potentials of unit depth, whose width is determined by k_{xy},

centred at $(0,0)$, (x_1,y_1), $(x_1,-y_1)$, $(-x_1,y_1)$ and $(-x_1,-y_1)$ where $x_1 = a_{0,\mathrm{S}}/2\sqrt{3}$ and $y_1 = a_{0,\mathrm{S}}/2$. As written above, $f(x_{\mathrm{ox}}\,,\,y_{\mathrm{ox}})$ is defined for the interval $-x_1 \leq x_{\mathrm{ox}} < x_1\,,\, -y_1 \leq y_{\mathrm{ox}} < y_1$ and is periodic such that:

$$f(x_{\mathrm{ox}} + 2x_1\,,\, y_{\mathrm{ox}} + 2y_1) = f(x_{\mathrm{ox}}\,,\, y_{\mathrm{ox}}) \tag{B.5}$$

The parameter δE_{frac} controls the barrier height and is defined as the fractional difference between the energy at one of the minima E_{min} and the energy at the top of the barrier between minima $E^{\ddagger}$:

$$\delta E_{\mathrm{frac}} = \frac{E^{\ddagger} - E_{\mathrm{min}}}{|E_{\mathrm{min}}|} \tag{B.6}$$

By choosing the width of the Morse potential sites in $f(x_{\mathrm{ox}}\,,\,y_{\mathrm{ox}})$ so that neighbouring sites only minimally interact, one can exert almost complete control over the barrier height. A schematic of the this surface potential is shown in Figure B.3.

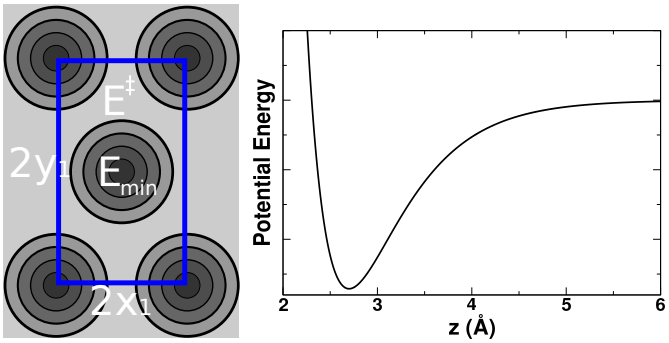

Figure B.3: Schematic of the external potential. The modulating function, $f(x_{\mathrm{ox}}\,,\,y_{\mathrm{ox}})$ is shown on the left and the Morse potential perpendicular to the surface is shown on the right. When $(x_{\mathrm{ox}}\,,\,y_{\mathrm{ox}})$ are at the centre of one of the symmetrised Morse functions shown on the left, the potential acting along z is left unchanged. For regions in between the symmetrised Morse functions, the z-component of the potential is increased up to a maximum value $E^{\ddagger}$, as determined by the parameter δE_{frac}.

The parameters used to define the external potential are given in Table B.2. The value for D_{z0} and δE_{frac} were taken from DFT calculations for a water monomer on Ag(111) that account accurately for the role of dispersion [60]. The width of the symmetrised Morse functions varied between surfaces such that the product of k_{xy} and the nearest neighbour distance remained constant, ensuring that the barrier height between minima on the different surfaces stayed the same.

Parameter	Value
D_{z0}	$27.112\,\mathrm{kJ/mol}$
k_z	$1.75\,\mathrm{\AA}^{-1}$
z_0	$2.7\,\mathrm{\AA}$
δE_{frac}	0.30605
$k_{xy} \times \frac{a_{0,\mathrm{S}}}{\sqrt{3}}$	18.0

Table B.2: Parameters used to define the external potential.

B.3 Assessment of the Water-Surface Interaction

Obviously, both of these surfaces are somewhat simplified when one considers the potentials exerted on water molecules by real surfaces and if one was attempting to accurately determine surface structures, sophisticated, system dependent, water–surface potentials would be required. Perhaps the most drastic simplification in these systems is that only the oxygen atoms are coupled to the surface and that the hydrogen atoms can vary their positions independently of the surface. This simplification is justified by noting that the traditional theory, to which we wish to compare, only makes reference to the relative positions of adsorption sites and does not account for any orientational dependence of the water molecules at the surface. In order to check that the lack of orientational dependence is not biasing the formation of non-bilayer overlayers, a GCMC simulation has been performed on the $\delta = 0.00$ HCP surface with a Morse potential acting on the z-component of the hydrogen atoms. A snapshot from this simulation is shown in Fig. B.4, where it can be seen that including this hydrogen term does not promote the formation of an ice-like bilayer. This hydrogen term was chosen arbitrarily, and the values are shown in Table B.3 (in general, it is not possible to decouple the hydrogen and oxygen interactions with the surface). However, by doing this, it is demonstrated that concerns regarding the hydrogen atoms approaching too close to the surface favouring non-bilayer structures is not the case with these simple potentials.

Furthermore, one can examine the results of previously reported DFT calculations on bilayers at Cu(110) and Ru(0001) (see the Reference [191] and its supporting information) in order to assess how important the hydrogen interaction with the surface is in influencing the oxygen corrugation and overall hydrogen

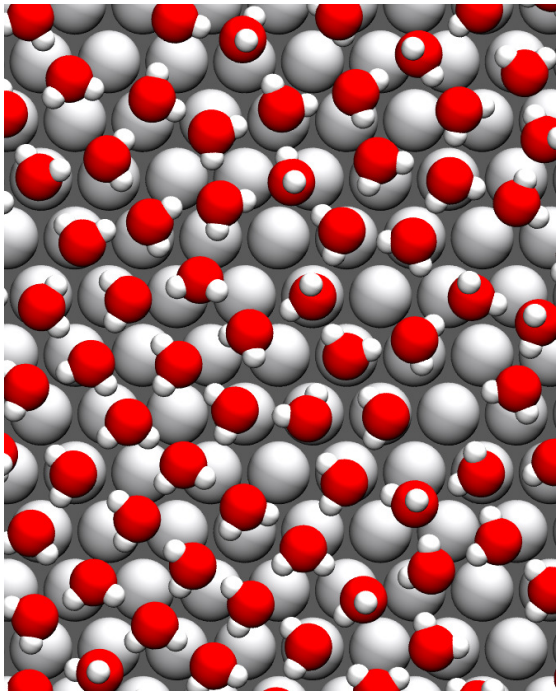

Figure B.4: Snapshot of a GCMC simulation on the HCP $\delta = 0.00$ surface with a Morse potential acting on the hydrogen atoms. It can be seen that inclusion of a potential of the hydrogen atoms (preventing close approach to the surface) does not promote ice-like bilayer formation.

Parameter	Value
D_{z0}, Oxygen	$17.112\,\mathrm{kJ/mol}$
D_{z0}, Hydrogen	$5.000\,\mathrm{kJ/mol}$
k_z, Oxygen	$1.75\,\mathrm{Å}^{-1}$
k_z, Hydrogen	$1.75\,\mathrm{Å}^{-1}$
z_0, Oxygen	$2.7\,\mathrm{Å}$
z_0, Hydrogen	$2.72\,\mathrm{Å}$
δE_{frac}, Oxygen	0.30605
$k_{xy} \times \frac{a_{0,\mathrm{S}}}{\sqrt{3}}$, Oxygen	18.0

Table B.3: Parameters used to define the external potential with hydrogen parameters. The choices for the hydrogen parameters have been chosen such that the total monomer binding energy was unchanged from the simulation set-up described in Table B.2. The potential on the hydrogen atoms only acts on the z-component of their positions.

bonding. The results from this DFT study (using the optB88-vdW density functional) are presented in Table B.4. Let us begin by comparing the H-down and H-up bilayers on Cu(110). For the H-down bilayer, it can be seen that there is very little difference in binding heights of the water oxygens ($\Delta z_{\mathrm{O}} = 0.055\,\mathrm{Å}$) between the flat lying water molecules and those water molecules with one of

their hydrogens pointed towards the surface, whereas in the case of the H-up bilayer, there is a far greater difference in binding heights ($\Delta z_{\mathrm{O}} = 0.776\,\text{Å}$). This increase in the buckling of the H-up bilayer is accompanied by an increase in the water-water bonding ($E_{\mathrm{gas}}^{\mathrm{H_2O-H_2O}} = -440$ meV/H_2O) relative to the H-down bilayer ($E_{\mathrm{gas}}^{\mathrm{H_2O-H_2O}} = -422$ meV/H_2O). However, the H-down bilayer, with its lower Δz_{O}, is overall more stable than the H-up bilayer and this is due to an increased binding with the surface ($E_{\mathrm{ads}}^{\mathrm{H_2O-Cu}} = -259$ and -178 meV/H_2O respectively). The findings that a buckling gives better water bonding, but that relaxation along the surface normal can outweigh this effect, are consistent with the conclusions in the main chapter (see Figure 3.4). If one looks at the bilayers on Ru(0001), we again see a similar trend, although the H-down and H-up bilayers have essentially identical adsorption energies.

		H-down	H-up
Cu(110)	E_{ads}	-681	-618
	$E_{\mathrm{gas}}^{\mathrm{H_2O-H_2O}}$	-422	-440
	$E_{\mathrm{ads}}^{\mathrm{H_2O-Cu}}$	-259	-178
	Δz_{O} (Å)	0.055	0.776
Ru(0001)	E_{ads}	-667	-666
	$E_{\mathrm{gas}}^{\mathrm{H_2O-H_2O}}$	-382	-447
	$E_{\mathrm{ads}}^{\mathrm{H_2O-Ru}}$	-285	-219
	Δz_{O} (Å)	0.516	0.771

Table B.4: Adsorption energies (E_{ads}) of different water overlayers and their decompositions on Cu(110) and Ru(0001). $E_{\mathrm{gas}}^{\mathrm{H_2O-H_2O}}$ is the energy of the fixed water layer (optimised at the surface) in the gas phase, $E_{\mathrm{ads}}^{\mathrm{H_2O-M}}$ is the binding energy due to water–metal interactions and Δz_{O} is the difference in heights of the water molecules above the surface.

In general, optimised H-down bilayers on metal surfaces tend to be flat relative to bulk ice (see e.g. References [45] and [232]), as is also the case on the hydroxylated kaolinite (001) surface [66]. This suggests that by only including the oxygen interaction with the surface, we are not ignoring an interaction that will structure the first layer into a highly corrugated bilayer. In fact, decomposition of DFT energies shows that stabilisation of corrugated water overlayers arises from augmentation of the hydrogen bonding interaction between water molecules, consistent with the presented findings, and not due to a repulsive interaction of the

hydrogen with the surface.

B.4 Water Binding Height Above Surface

As mentioned in the main text, the water molecules all bind at near equal heights above the surface. Figure B.5 shows a histogram of water oxygen heights above the HCP surface and snapshot from the honeycomb surface to demonstrate this fact.

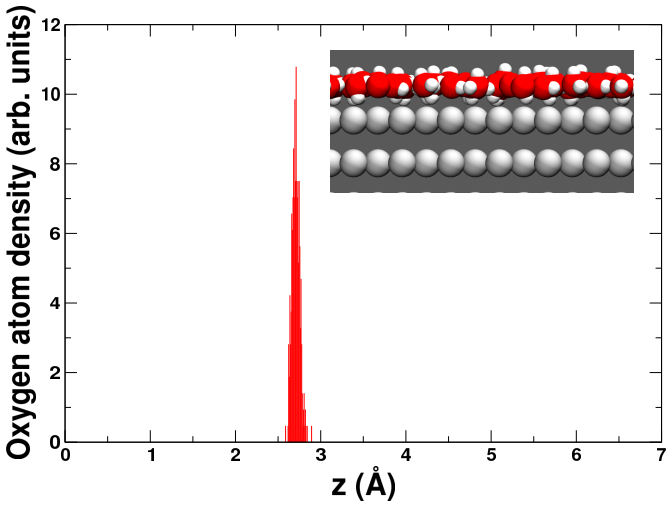

Figure B.5: Histogram of water oxygen heights above one of the HCP surfaces. All water molecules are bound within 0.5 Å of each other. The inset shows a snapshot from one of the honeycomb surfaces, showing that the water molecules bind at near equal heights. The data shown is representative of all surfaces and therefore for clarity, data from the other surfaces is not shown.

B.5 Results from the Honeycomb Surface

Figure B.6 shows the RDFs on the honeycomb surfaces. The results are qualitatively similar to those on the HCP surfaces. One can compare the energetics of the different overlayers by varying the depth of the Lennard-Jones potential on each surface until the binding energy roughly matches that of the $\delta = 0.00$ surface and performing a new set of canonical MC simulations without the need for a new set of GCMC simulations. The binding energies for the new surfaces are given in Table B.5. From Figure B.7 it can be seen that the contribution to the total energy from the water–water interactions remains effectively constant on all surfaces, that is, there is no significant energy cost in forming disordered structures on the compressed surfaces. It can also be seen that the total energy is

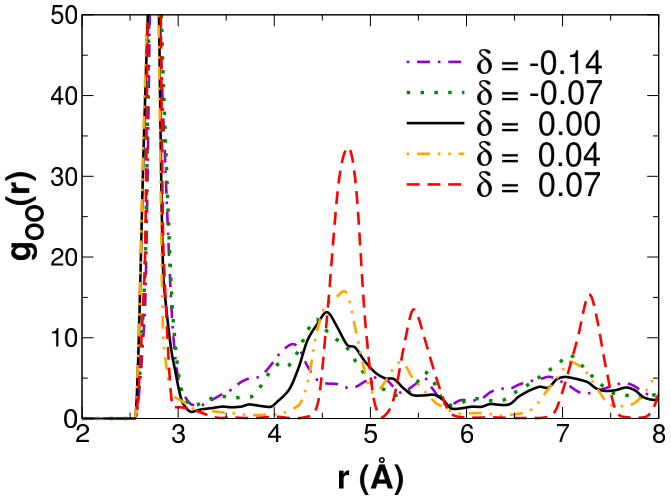

Figure B.6: Radial distribution functions for different δ (honeycomb surface). As on the HCP surfaces, the position of the nearest–neighbour peak is essentially invariant as the substrate lattice constant is changed, located at approximately 2.75 Å.

δ	ϵ (kJ/mol)	E_{bind} (kJ/mol)
−0.14	6.081	−28.64
−0.07	6.584	−28.65
0.00	6.991	−28.64
0.04	7.036	−28.62
0.07	7.321	−28.65

Table B.5: Altered Lennard-Jones depths and new binding energies. The $\delta = 0.00$ remained unchanged. The width of the LJ potential ($\sigma = 3.346$ Å) was left unchanged.

reasonably flat with the disregistry, with a slight stabilisation going from $\delta = 0.04$ to $\delta = 0.07$. As the water-water interactions are similar across-the-board, it is concluded that this stabilisation arises from increased binding to the surface. This result is not surprising given the similarity between $r_{0,\mathrm{i}}$ (2.75 Å) and $r_{0,\mathrm{S}}$ on the $\delta = 0.07$ surface (2.74 Å).

B.6 Results Using SPC/E Water

To test how dependent the results are on the choice of water model, the same set of simulations have been performed using SPC/E water. Before presenting the results, it should be noted that the SPC/E water dimer is even more tightly binding than the TIP4P dimer (the dimer binding energy of SPC/E is −30.1 kJ/mol) [233]. This means that although the water–water and water–surface interactions are still comparable, the water–water interactions are slightly stronger, in contrast

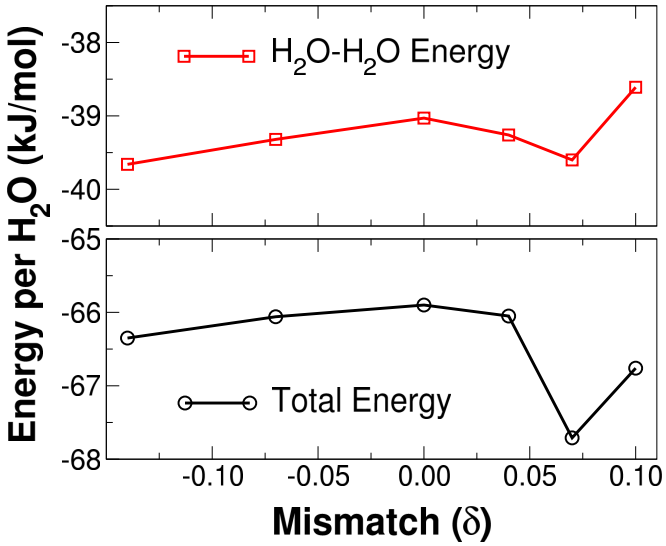

Figure B.7: Energies of the different overlayers on the honeycomb surface. The water–water energy was calculated by removing the surface from the final configuration from the canonical MC runs and performing single point energy calculations on the remaining water structures. The figure shows that there is very little change in water–water interaction for the different surfaces. The total energy was calculated from averages over the canonical MC runs. Again, there is little change in energy from changing $a_{0,S}$ except on the $\delta = 0.07$ surface where there is a slight stabilisation. As the water–water energy remains almost constant, it is concluded that this is due to an increase in the water–surface interaction.

to the TIP4P simulations. The GCMC simulations on the honeycomb surfaces were performed using the parameters that gave near equal binding on all surfaces (see Table B.5). Figures B.7 and B.9 show the energy profiles using SPC/E water. The agreement between the two water models is good, with some small differences on the honeycomb surfaces with compressed δ, probably owing to the differences in water–water binding.

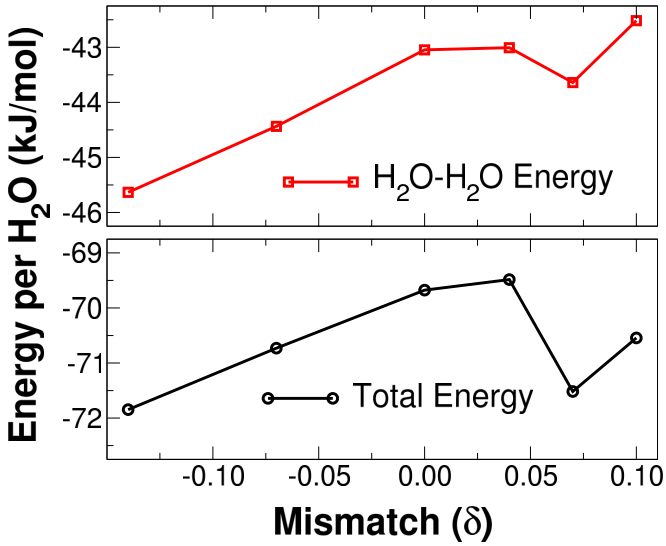

Figure B.8: Energies of the different overlayers on the honeycomb surfaces using SPC/E water. On the whole we see a similar trend to TIP4P, except that the total energy at compressed δ is relatively more stable. As we can see, this arises from an increase in the water–water interaction, which may be expected due to tighter binding of SPC/E water.

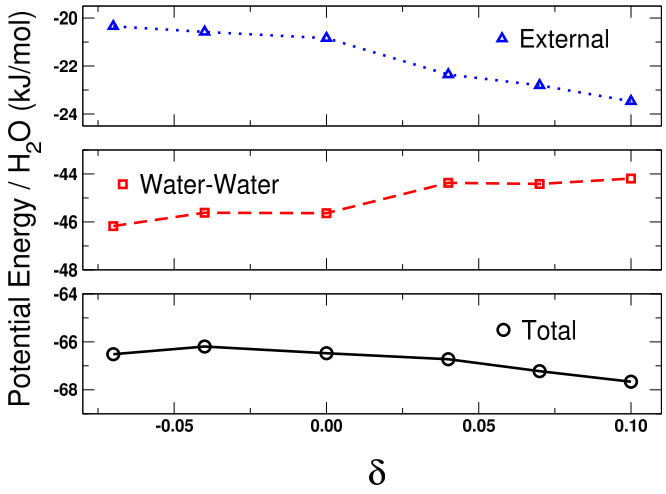

Figure B.9: Energies of the different overlayers on the HCP surface using SPC/E water. The agreement with the TIP4P simulations is very good (c.f. Figure 3.3).

Appendix C

Supplementary Material for Chapter 4

C.1 Overview

This appendix provides supplementary material for Chapter 4. This includes ambient temperature density profiles, a committor analysis, and details of the adsorption energy calculations. Movies are provided on the enclosed CD-ROM and online [103].

C.2 Movies of Heterogeneous Ice Nucleation

C.2.1 Movie 1

In this movie, the nucleation event shown in Figures 4.3 and 4.4 (b) is presented. In this trajectory, the structuring of the second layer before the first can clearly be seen.

C.2.2 Movie 2

This movie shows the nucleation event shown in Figure 4.4 (d). In this instance, the structuring of the first and second layers happens in a more concerted fashion than for the trajectory presented in Movie 1. In Section C.5, however, it will be seen that the density changes occurring away from the surface in Figure 4.4 (d) are part of the nucleation mechanism and are not merely indicative of growth.

C.3 Movies Showing how the Density Normal to the Surface Changes During Nucleation

C.3.1 Movie 3

This movie shows how the density profile changes in time for the trajectory shown in Movie 1. The densities are shown as running averages over a 2.5 ns interval. The density changes away from the surface clearly occur before those occurring at the surface.

C.3.2 Movie 4

Here it is shown how the density profile changes in time for the trajectory shown in Movie 2. At approx. 278 ns there is a concerted change in the density in the first, second and third layers. Structuring of the layers then appears to occur first in the first layer, followed very quickly by the second layer. The densities are shown as running averages over a 2.5 ns interval.

C.4 Density Profiles at Ambient Temperature

The high density of liquid water at the hydroxyl-terminated (001) face of kaolinite is in fact a feature at ambient temperatures and not a consequence of the water being supercooled. This is shown in Figure C.1, where the density of water perpendicular to the kaolinite surface at 220 K and 300 K is shown (averaged over 1 ns time interval). The water density perpendicular to the kaolinite surface has also been calculated using the PBE density functional [113] with 132 water molecules over a 15 ps time interval at 330 K, and also shows a high density peak at the surface. The DFT simulations were performed by Dr. Xiaoliang Hu using the `CP2K` simulation package [234]. Orbitals were described by an atom centred Gaussian-type basis set while an auxiliary plane wave basis set was used to re-expand the electron density [235]. Analytic Goedecker-Teter-Hutter pseudopotentials [236, 237] were employed to represent the core electrons. For the valence electrons the basis sets used consisted of short ranged quadruple-ζ (hydrogen and oxygen atoms), triple-ζ (aluminium) and double-ζ (silicon) basis functions with two sets of polarisation functions, except for silicon, where only one set of polari-

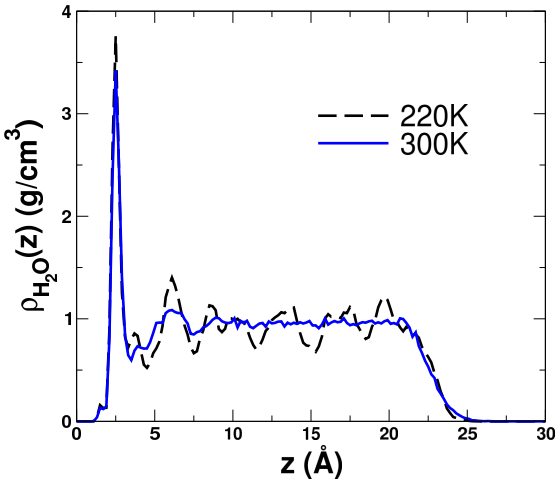

Figure C.1: Comparison of water densities perpendicular to the kaolinite slab at 220 K and 300 K. The profiles shown are averages over a 1 ns time interval. Although the supercooled liquid is overall more structured than at 300 K, the high density peak at the surface still features at the higher temperature.

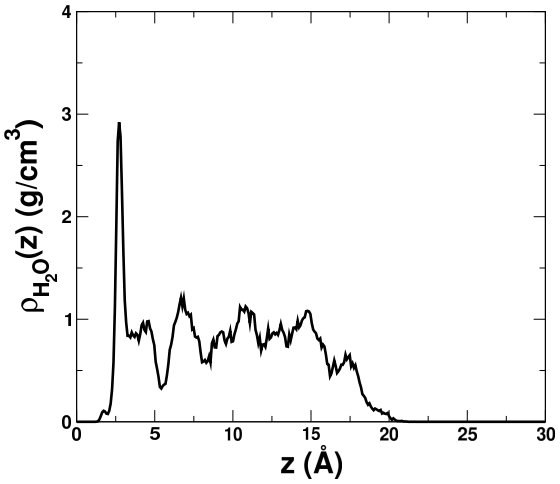

Figure C.2: Density profile of water perpendicular to the kaolinite slab using the PBE density functional at 330 K. The profile is an average over 15 ps with 132 water molecules. The density is overall more structured than that obtained from the 300 K force field simulation, but the high density at the surface relative to the rest of the system is still observed. Courtesy of Dr. Xiaoliang Hu.

sation functions was used. The plane wave cutoff for the electron density was set to 280 Ry. The Brillioun zone was sampled at the Γ-point only. The convergence criterion for the wavefunction optimization was set to 1.0×10^{-7} Ha between final self-consistent field iterations. The mass of the hydrogen atoms was set to that of deuterium and a 1 fs timestep was used to propagate the dynamics, with a Nosé-Hoover chain of length 4 used to maintain the target temperature of 330 K.

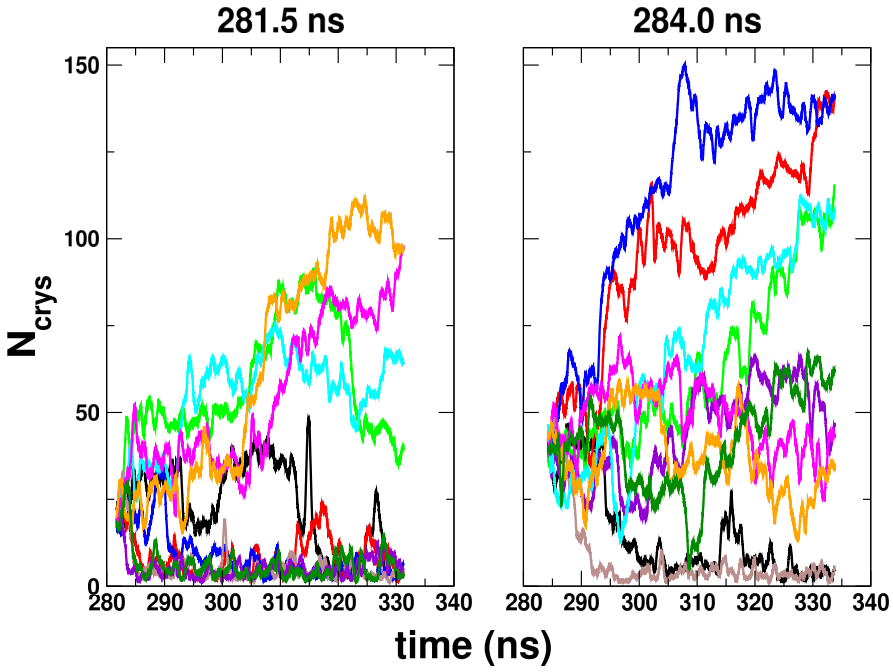

Figure C.3: Committor analysis from one of the heterogeneous ice nucleation trajectories. Results here are shown for initial configurations sampled at 281.5 ns and 284.0 ns from the initial ice forming trajectory. It is clear that by 284.0 ns the initial trajectory has not entered a post-critical regime.

C.5 Committor Analysis for Trajectory 'd' of Figure 4.4

In Figure C.3 the preliminary results from a committor analysis for trajectory 'd' of Figure 4.4 are presented, using the same method as described in the main text. Here configurations have been used from two times, 281.5 ns and 284.0 ns, of the original trajectory as initial configurations for the analysis. According to this analysis, neither of the configurations can be deemed post-critical. As the density profile in Figure 4.4 (d) of the main paper is measured at 282.5 ns of the main trajectory, one has confidence that the observed density changes are happening during the nucleation event itself, and are not indicative of growth of the ice crystal.

C.6 Adsorption Energies of the Prism and Basal Faces of Ice to Kaolinite (001)

Structures for the prism face of ice bound to kaolinite were taken directly from a snapshot of one of the ice forming trajectories. For the basal face bound to kaolinite, structures were built based on the results obtained for the fist contact layer in Reference [66], using the same size cell as those used in the MD simulations. The DFT geometry optimisations were performed using the same initial structures as the force field calculations. The definition of the adsorption energy per water molecule is:

$$E_{\text{ads}} = \frac{(E_{\text{water/slab}} - E_{\text{slab}} - N_{\text{water}} E_{\text{water}})}{N_{\text{water}}} \tag{C.1}$$

where $E_{\text{water/slab}}$ is the total energy of the water molecules adsorbed at the surface, E_{slab} is the energy of the bare slab, E_{water} is the energy of the isolated water monomer and N_{water} is the number of water molecules. As the structures for the prism face were extracted from a finite temperature MD simulation, the energy of the bare slab was calculated separately for each structure. The DFT calculations were performed using the `CP2K` simulation package with the PBE density functional. Orbitals were described by an atom centred Gaussian-type basis set while an auxiliary plane wave basis set was used to re-expand the electron density. Analytic Goedecker-Teter-Hutter pseudopotentials were employed to represent the core electrons. For the valence electrons the basis sets used consisted of short ranged molecularly optimised double-ζ basis functions with a single set of polarization functions (DZVP). It has been shown that such basis set functions greatly reduce the effects of basis set superposition error [238]. The plane wave cutoff for the electron density was set to 400 Ry. The Brillioun zone was sampled at the Γ-point only. The convergence criterion for the wavefunction optimisation was set to 1.0×10^{-6} Ha between final self-consistent field iterations. Geometries were converged when the forces on all atoms were below 5.0×10^{-5} Ha/Bohr. For all calculations a minimum vacuum gap of 15 Å between periodic images was used throughout. The calculations were also performed using a regular DZVP

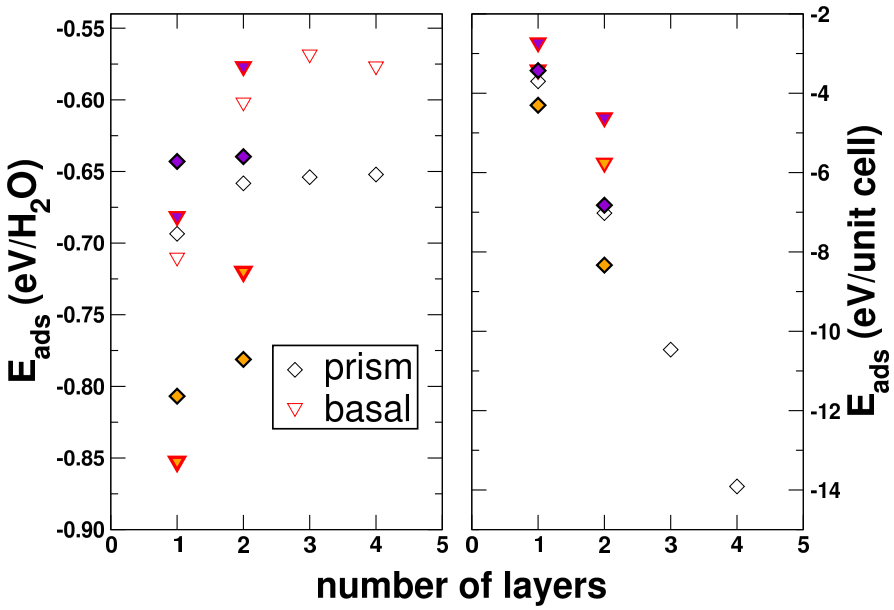

Figure C.4: Adsorption energy of ice bound to kaolinite through its prism and basal faces. The empty symbols are results using the force field (same as Fig. 5 in the main article). The results shown by the purple filled symbols have been obtained using the molecularly optimised DZVP basis set, whereas the yellow filled symbols show results obtained with the regular DZVP basis set. Although there is a large difference in adsorption energies between the two basis sets, the trend that the prism face becomes more stable upon adsorption of the second water layer is found with both.

basis set, which despite drastically changing the adsorption energies, produced the same trends observed with the molecularly optimised DZVP basis set. The DFT results presented in Figure 4.6 were obtained with the molecularly optimised orbitals. In Figure C.4 results using both the DZVP and molecularly optimised DZVP basis sets are shown.

For the liquid layer calculations, 25 configurations were randomly chosen from different trajectories. Molecules whose oxygen atom position was greater than 5 Å above the average height of the surface oxygen atoms were then subsequently removed (this corresponds to a distance from the surface that approximately in the middle between the first and second ice peaks). The remaining water molecules and surface were then allowed to relax. The error bars in Figure 4.6 correspond to the standard deviation in the adsorption energy.

Appendix D

Nucleation Mechanism on Graphene Nanoflakes

In this appendix, the nucleation mechanism on the GNFs reported in Chapter 5 is discussed in more detail. Let us begin by noting that, for the GNFs with $0.1 < E_{\text{ads}}/\Delta H_{\text{vap}} < 1.4$, a layering of interfacial water molecules that is similar to that reported by Lupi *et al.* [129, 130] is observed, as shown in Figure D.1 (a) for $E_{\text{ads}}/\Delta H_{\text{vap}} \approx 0.16$. A layering at the GNFs that bind water more strongly is also seen, shown for example in Figure D.1 (b) for $E_{\text{ads}}/\Delta H_{\text{vap}} \approx 1.9$. In the case of the latter, there is a strong, sharp peak that does not change appreciably after ice nucleation has occurred (in contrast, for $E_{\text{ads}}/\Delta H_{\text{vap}} \approx 0.16$, the first density peak splits upon ice nucleation, indicative of ice-like bilayer formation). In References [129] and [130] it was this layering that was used to explain the ice nucleating ability of the GNF. Although Lupi *et al.* noted that layering also occurs on other good ice nucleating agents [20, 239], the generality of the importance of layering was left as an open question. Figure D.2 compares the density profiles of water above the two GNFs presented in Figure D.1 with those obtained above the FCC-111 NP for $E_{\text{ads}}/\Delta H_{\text{vap}} \approx 0.08$ and 1.9, and with the GNF for $E_{\text{ads}}/\Delta H_{\text{vap}} \approx 0.08$. These three particles yield rates that are slower than or comparable to homogeneous nucleation. Apart from some subtle differences, the layering exhibited above surfaces of similar E_{ads} are broadly the same, suggesting that the layering of interfacial water on its own is insufficient to promote ice

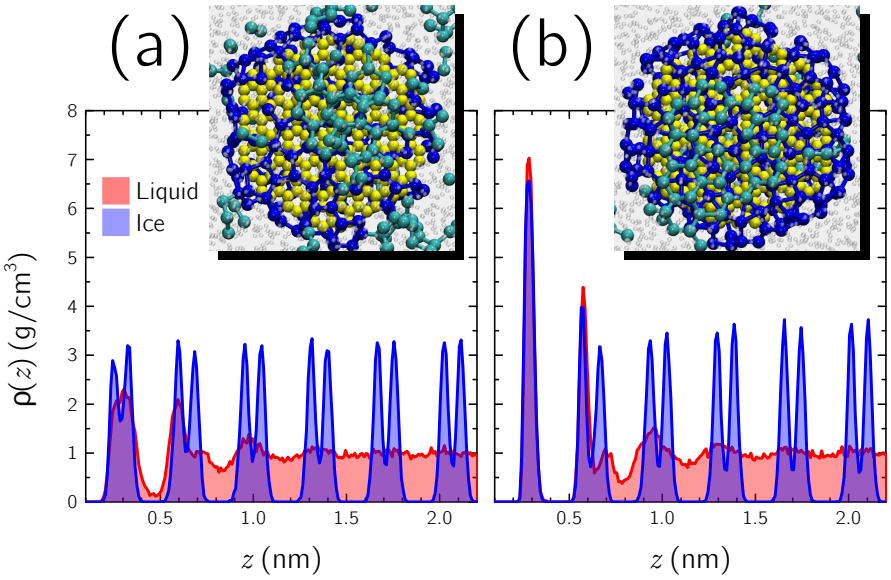

Figure D.1: Density profiles $\rho(z)$ of liquid water (red) and ice (blue) at a height z above the GNF for (a) $E_{\text{ads}}/\Delta H_{\text{vap}} \approx 0.16$ and (b) $E_{\text{ads}}/\Delta H_{\text{vap}} \approx 1.9$. In both cases a significant layering of the interfacial water is seen. In (a), the splitting of the first peak after ice nucleation is indicative of an ice-like bilayer forming at the surface. In (b), the first peak is much sharper and remains essentially unchanged after nucleation. Apart from this difference in the contact layer, the nucleation mechanism appears very similar between these two GNFs.

nucleation.

Here it will be argued that the GNFs act to enhance nucleation by acting like smooth structureless walls, in contrast to the FCC-111 NPs. The first piece of evidence to support this claim is that the structure of the contact layer for $E_{\text{ads}}/\Delta H_{\text{vap}} > 1.4$ is strikingly similar to the "wheels" quasicrystal obtained in Reference [240] upon confining water between two structureless walls under high pressure. An enlarged image of such a structure is shown in Figure D.3. In Reference [240] it was also shown that with increasing pressure, confined water undergoes a transition from forming hexagonal ice-like layers upon freezing to forming the quasicrystal, which is the same behaviour we see at the GNFs upon increasing E_{ads}. It therefore appears that water at the GNFs is behaving in an analogous manner to water confined between two structureless walls. With regard to the enhanced nucleation rate seen with the GNFs, the freezing temperatures for these confined layers of water (approximately 290 K at zero lateral pressure[1], falling to 235 K at 4000 bar) are significantly higher than the freezing temperature

[1]The equilibrium melting temperature for confined bilayer water is actually higher than that of bulk water. This is discussed in detail in Reference [241].

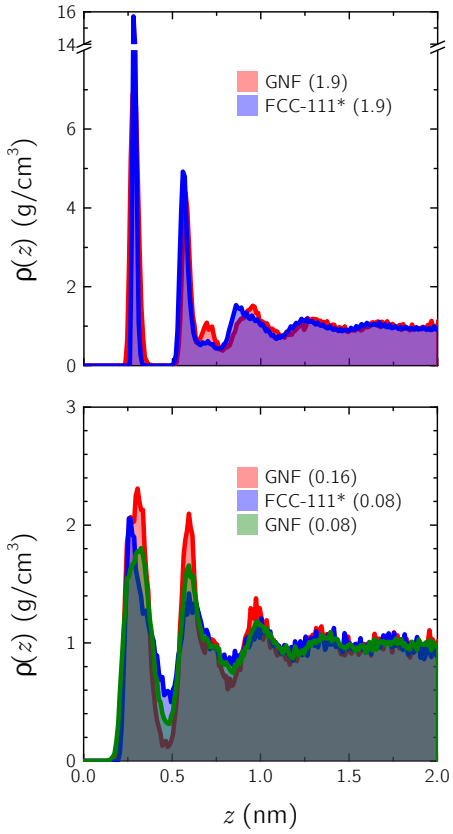

Figure D.2: Density profiles of liquid water $\rho(z)$ at a height z above different nanoparticle surfaces (the value of $E_{\text{ads}}/\Delta H_{\text{vap}}$ is shown in parentheses). GNF (1.9) and GNF (0.16) both enhance ice nucleation, whereas the others either inhibit or have little effect on the rate. For surfaces of comparable E_{ads}, $\rho(z)$ is broadly similar, suggesting that layering of interfacial water on its own is insufficient to enhance ice nucleation. As it is only relative peak positions that matter, for ease of comparison, FCC-111 (0.08) and FCC-111 (1.9) have been offset by +0.075 and +0.04 nm, respectively: this is indicated by the asterisk (*) in the legend.

of bulk homogeneous water of 202 K [121], and it is therefore reasonable that they should undergo nucleation on a shorter timescale than the bulk liquid.

Of course, given the similarity of the density profiles above the GNF and FCC-111 NPs at high E_{ads}, it would also seem reasonable to argue that the interfacial water at the FCC-111 NP should also behave like confined water and freeze more easily. To reconcile this, the probability density $P(x, y)$ of water molecules in the plane of the surface prior to nucleation[2] has been computed, such that

[2]For the FCC-111 NP, $P(x, y)$ was obtained from 25 ns of a simulation prior to nucleation. For the GNFs, 12 ns and 14 ns of a trajectory were used for $E_{\text{ads}}/\Delta H_{\text{vap}} \approx 0.16$ and 1.9, repsectively. At the GNF, a water molecule is defined as being in the contact layer if it is a height $0 < z < 0.45$ nm from the surface, and the second layer if $0.45 < z < 0.8$ nm. For the FCC-111, the criteria are $0 < z < 0.375$ nm and $0.375 < z < 0.725$ nm for the contact and second layers, respectively.

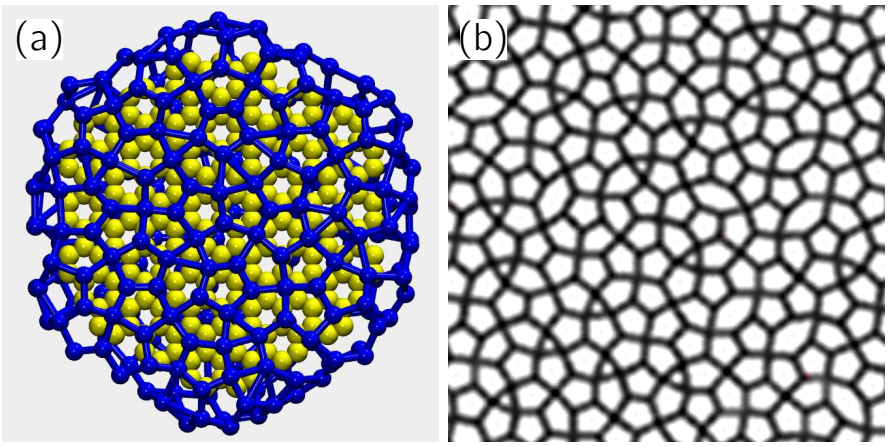

Figure D.3: In (a), molecules in the contact layer at the GNF for $E_{\text{ads}}/\Delta H_{\text{vap}} > 1.4$ are shown. Panel (b) is adapted from Reference [240] and shows the "wheels" quasicrystal formed by water confined between flat structureless surfaces under pressure. The molecules at the strongly adsorbing GNF bear a striking resemblance to the "wheels" quasicrystal. Carbon atoms are shown in yellow, and water molecules in the contact layer in blue.

$P(x, y)\,\mathrm{d}x\mathrm{d}y$ is the probability of finding a water molecule in an area element $\mathrm{d}x\mathrm{d}y$. From $P(x, y)$ one can then define the in-plane free energy for a water molecule as:

$$\beta\Delta f(x, y) = -\ln\left[P(x, y)\mathrm{d}x\mathrm{d}y\right]. \tag{D.1}$$

This is shown in Figure D.4. Despite the similarities of density profiles of water above the GNF and FCC-111 NP for $E_{\text{ads}}/\Delta H_{\text{vap}} \approx 1.9$, it is clear that the structure of water molecules within the layers is distinctly different. Importantly, this is true not only for the contact layer, but also for the second layer. Specifically, at the GNF it is seen that, the although the quasicrystal-like structure forms at the surface, the water molecules do not occupy distinct adsorption sites as they do at the FCC-111 NP (in fact, the quasicrystal has a diffusivity between liquid and ice [240] meaning significant structural rearrangements are observed). Consequently, no obvious in-plane structure is observed in the second layer. On the other hand, at the FCC-111 NP it is seen that the water molecules in the second layer have an obvious preference to be directly above the water molecules in the contact layer. As the structure in the contact layer at this high value of E_{ads} does not resemble ice, this has the effect of promoting a frustrated structure in the second layer. It is concluded that for layering of interfacial water to promote ice nucleation, $\beta\Delta f(x, y)$ must be relatively smooth so that the water molecules can rearrange

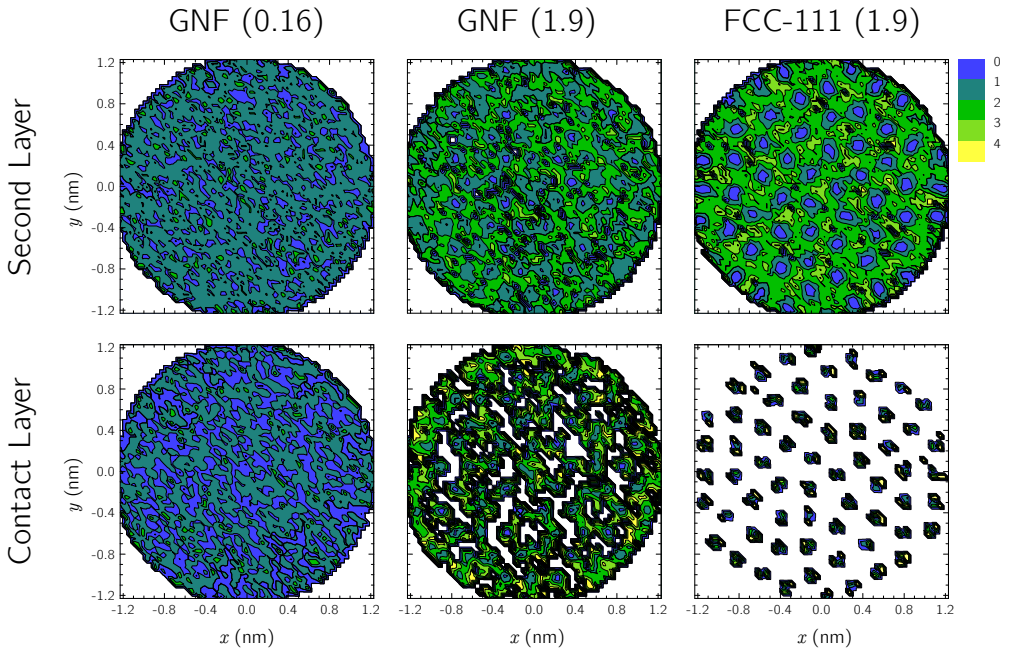

Figure D.4: In-plane free energy surface $\beta\Delta f(x, y)$ for a water molecule at the FCC-111 NP and two GNFs (the numbers in parentheses indicate the value of $E_{\text{ads}}/\Delta H_{\text{vap}}$). The lower and upper panels show results for the contact and second layers, respectively. At FCC-111 (1.9), the structure of the contact layer has a marked effect on the structure of the second layer. At GNF (1.9), although the quasicrystal-like structure forms, the molecules in the contact layer do not adsorb at specific sites on the surface and consequently, $\beta\Delta f(x, y)$ in the second layer at the GNF is a smoother function than at the FCC-111 (1.9). For GNF (0.16), $\beta\Delta f(x, y)$ is smooth for both the contact and second layers.

into an ice-like configuration. For surfaces that promote adsorption at specific sites, such as the FCC-111, the layering of interfacial water does not promote ice nucleation.

Appendix E

Supplementary Material for Chapter 5

E.1 Overview

This appendix provides supplementary material for Chapter 5. In particular, details of the fitting procedure to extract the nucleation rate are provided and time-resolved snapshots of selected nucleation events are presented. Movies of these nucleation events are provided on the enclosed CD-ROM.

E.2 Fitting procedure

For each system, 16 simulations were performed. As nucleation is a stochastic process, even in the same system, one has to wait for different times to see nucleation. One can determine the induction time to nucleation t_{ind} for each trajectory by monitoring the time evolution of the potential energy and fitting to the equation:

$$U(t) = U_0 + \frac{\Delta U}{1 + \exp(k(t - t_{\mathrm{ind}}))}. \tag{E.1}$$

In Equation E.1, U_0, ΔU, k and t_{ind} are all freely variable parameters. See Figure E.1 (a) for an example. From the distribution of induction times to nucleation, for each system it is then possible to calculate the probability that a simulation

will be in the liquid state at a time t after the simulation has started (at $t = 0$):

$$P_{\mathrm{liq}}(t) = 1 - \frac{1}{N_{\mathrm{sim}}} \sum_{i=1}^{N_{\mathrm{sim}}} \Theta(t - t_{\mathrm{ind}}^{(i)}), \tag{E.2}$$

where $N_{\mathrm{sim}} = 16$ is the total number of simulations performed for each system, $t_{\mathrm{ind}}^{(i)}$ is the induction time determined for the i^{th} simulation, and $\Theta(x)$ is the heaviside step function. Typical $P_{\mathrm{liq}}(t)$ data are shown in Figure E.1 (b). From the shape of these curves, one can quantify a nucleation rate R for each system by fitting:

$$P_{\mathrm{liq}}(t) = \exp\left[-(Rt)^{\gamma}\right], \tag{E.3}$$

where $\gamma > 0$ is also a fitting parameter. The reason for using this form of the fitting function rather than the regular exponential function (i.e. with $\gamma = 1$) is because the nucleation simulations start from a non-equilibrium distribution of initial phase space points. This means that the system needs time to relax towards its equilibrium state. When nucleation is fast, for example in the presence of the FCC-111 NP for $E_{\mathrm{ads}}/\Delta H_{\mathrm{vap}} \approx 0.4$, this relaxation time and the time required for nucleation become comparable, and non-exponential kinetics is observed. Examples are given in Figure E.2. In these cases, compressed exponential kinetics ($\gamma > 1$) are actually found. Such behaviour has also been reported in protein folding models [242].

The actual fits were obtained using version 8.6 of the `OriginPro` software package from OriginLab, Northampton, MA. `OriginPro` uses the Levenberg-Marquardt algorithm to adjust the parameter values in an iterative procedure. It also reports standard errors S_R and S_γ in the fitted parameters R and γ, respectively, by calculating the square root of the diagonal elements of the variance-covariance matrix. The error in the ratio is then calculated as:

$$S_{R/R_{\mathrm{hom}}} = \frac{R}{R_{\mathrm{hom}}} \sqrt{\left(\frac{S_R}{R}\right)^2 + \left(\frac{S_{R_{\mathrm{hom}}}}{R_{\mathrm{hom}}}\right)^2}, \tag{E.4}$$

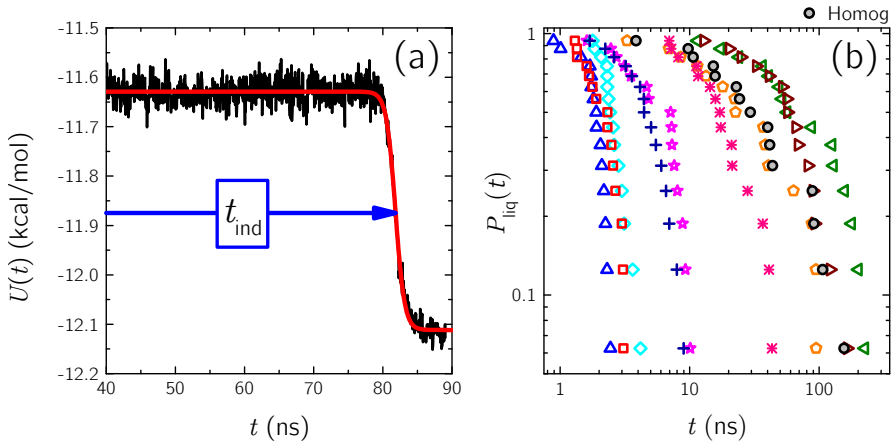

Figure E.1: In (a) an example is shown of how the induction time to nucleation t_{ind} for a trajectory is determined. The black line shows the potential energy extracted from the simulation, and the red line shows a fit to these data using Equation E.1. For each system a distribution of t_{ind} is obtained, and by Equation E.2 one can calculate $P_{\mathrm{liq}}(t)$, shown in (b). Data for bulk homogeneous nucleation is shown by filled circles. The other data show typical $P_{\mathrm{liq}}(t)$ curves obtained in the presence of the FCC-111 NP (different symbols correspond to different values of E_{ads}).

and the error in the logarithm is computed as:

$$S_{\log_{10}} = \frac{S_{R/R_{\mathrm{hom}}}}{\frac{R}{R_{\mathrm{hom}}}\ln(10)}. \tag{E.5}$$

It is this value that is indicated by the error bars in Figure 5.1.

E.3 Snapshots of Nucleation Events

In Figures E.3, E.4 and E.5 snapshots from three nucleation events in the presence of FCC-111 (0.90), GNF (0.16) and smooth (0.56) are shown – the numbers in parentheses indicate values of $E_{\mathrm{ads}}/\Delta H_{\mathrm{vap}}$. For the FCC-111 NP, it has been decided to present a sub-optimal value of $E_{\mathrm{ads}}/\Delta H_{\mathrm{vap}}$ as it nicely demonstrates the competition between adsorption at the excess binding sites and the in-plane template effect. For optimal values of $E_{\mathrm{ads}}/\Delta H_{\mathrm{vap}}$, the water molecules in the contact layer form an hexagonal overlayer almost immediately. In all three figures, the colour scheme is the same as that used in Figure 5.2.

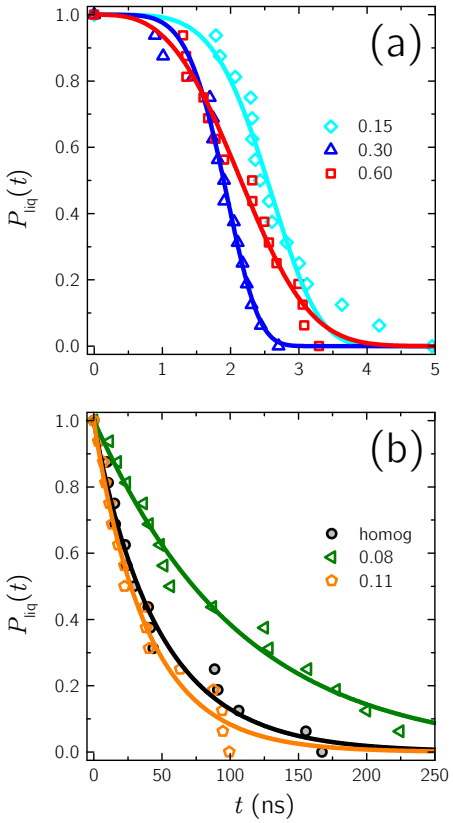

Figure E.2: In (a) $P_{\text{liq}}(t)$ data is shown for the FCC-111 NP with $E_{\text{ads}}/\Delta H_{\text{vap}} \approx 0.15$, 0.30 and 0.60, and their corresponding fits using Equation E.3. For these systems nucleation is fast and the $P_{\text{liq}}(t)$ is clearly non-exponential, which is reflected by the obtained fitted values of $\gamma = 4.79$, 5.25 and 3.08, respectively. In (b), $P_{\text{liq}}(t)$ is shown for $E_{\text{ads}}/\Delta H_{\text{vap}} = 0.08$ and 0.11 and for the bulk homogeneous system. In these systems, nucleation is a much slower process than relaxation from the initial state, and exponential kinetics is observed. In these cases, $\gamma = 0.96$, 1.02 and 0.98 for the homogeneous and $E_{\text{ads}}/\Delta H_{\text{vap}} = 0.08$ and 0.11 systems, respectively.

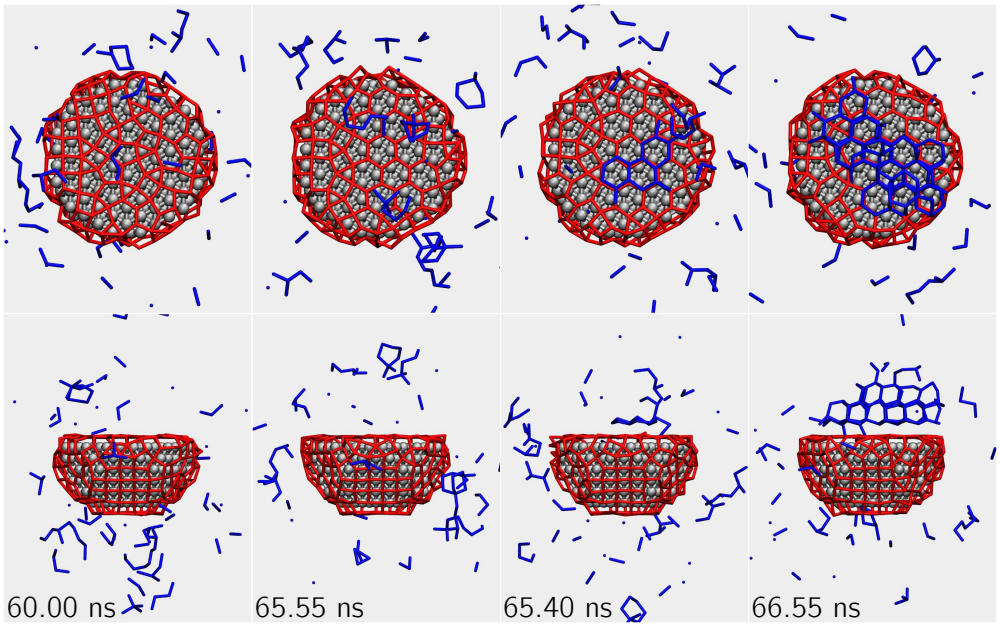

Figure E.3: Nucleation in the presence of the FCC-111 NP with $E_{\text{ads}}/\Delta H_{\text{vap}} \approx 0.90$. For this sub-optimal value of $E_{\text{ads}}/\Delta H_{\text{vap}}$, an area of decreased coverage must occur at the (111) surface in order to lead to an hexagonal arrangement of water molecules that can act as a template for the water molecules in the layers above. See also Movie 1.

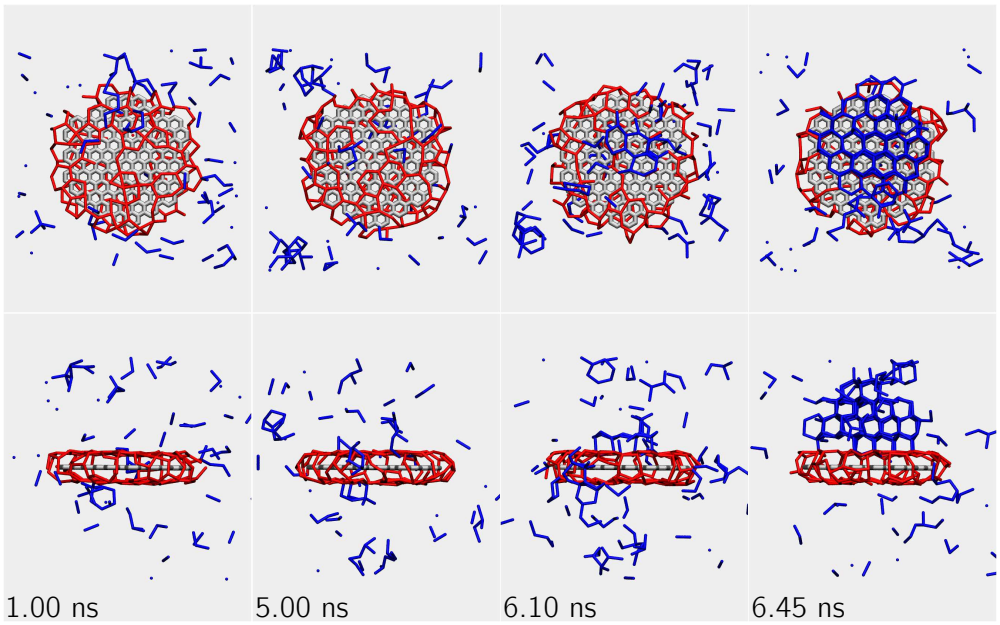

Figure E.4: Nucleation in the presence of the GNF with $E_{\text{ads}}/\Delta H_{\text{vap}} \approx 0.16$. Although an hexagonal layer forms at the GNF upon ice nucleation, there is no in-plane templating by the surface (the ice is incommensurate with the substrate). Ice formation at the surface and in the water layers above occurs simultaneously. See also Movie 2.

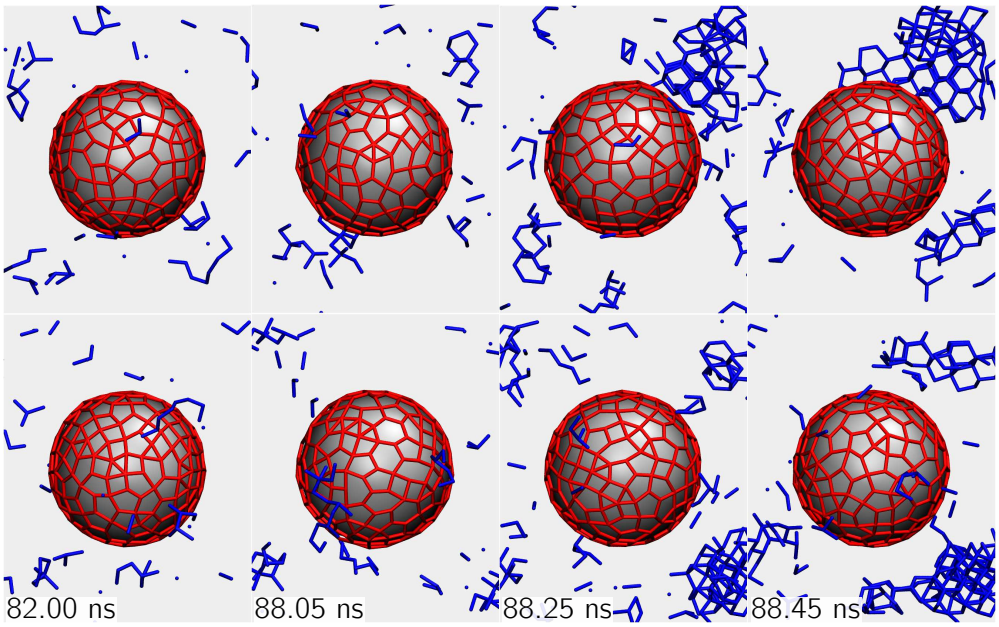

Figure E.5: Nucleation in the presence of the smooth NP with $E_{\text{ads}}/\Delta H_{\text{vap}} \approx 0.56$. The top and bottom panels are the same snapshot from different angles. Nucleation occurs away from the NP and the structure of the water in the contact layer is unaffected. See also Movie 3.

Appendix F

Supplementary Material for Chapter 6

F.1 Overview

This appendix provides supplementary material for Chapter 6. In particular, a relationship between the path free energy (given by Equation 6.12) and the free energy derived from equilibrium statistical mechanics is derived. Details of the q_{12} order parameter are also given, and results with 1728 mW molecules are presented to support the conclusions of the main chapter.

F.2 From Path to Equilibrium Free Energies

In this part of the appendix it is shown that the path free energy given by Equation 6.12 is equivalent to the 'regular' free energy of equilibrium statistical mechanics. In what follows, distributions relating to the path ensemble are denoted with a tilde, whereas regular statistical distributions are written without the tilde. This derivation assumes deterministic trajectories.

Let us begin by stating that for deterministic trajectories, $x(t)$, the distribution functional is given by:

$$\tilde{\rho}[x(t)] = \rho(\mathbf{x}_0) \prod_{0<t'\leq t} \delta[x_{t'} - x_{t'}(\mathbf{x}_0))], \tag{F.1}$$

where $\rho(x_0)$ is the distribution functional of initial phase space points:

$$\rho(\mathbf{x}_0) \propto \frac{\exp\left[-\beta H(\mathbf{x}_0)\right]}{Q}, \tag{F.2}$$

where Q is the appropriate partition function and $H(\mathbf{x}_0) \equiv H(\mathbf{r}_0^N, \mathbf{p}_0^N)$ is the Hamiltonian of the system. If we consider an order parameter $c[x(t)]$ of a trajectory $x(t)$ then this will be distributed according to:

$$\begin{aligned}
\tilde{P}(c) &= \sum_{x(t)} \delta(c[x(t)] - c)\tilde{\rho}[x(t)] \\
&= \sum_{\mathbf{x}_0} \delta(c[x(t);\mathbf{x}_0] - c)\tilde{\rho}[x(t)] \\
&= \frac{\int \mathrm{d}\mathbf{r}_0^N \mathrm{d}\mathbf{p}_0^N \delta(c[x(t);\mathbf{r}_0^N,\mathbf{p}_0^N] - c)\exp[-\beta H(\mathbf{r}_0^N,\mathbf{p}_0^N)]}{Q}.
\end{aligned} \tag{F.3}$$

In the second line, the notation $c[x(t);\mathbf{x}_0]$ has been introduced to indicate that for deterministic trajectories, the value of $c[x(t)]$ can be determined solely by the initial phase space point of the trajectory.[1] It is thus clear that c can act as an order parameter for initial phase space points. One can therefore write a reduced partition function:

$$Q(c[x(t);\mathbf{r}_0^N,\mathbf{p}_0^N]) = \int \mathrm{d}\mathbf{r}_0^N \mathrm{d}\mathbf{p}_0^N \delta(c[x(t);\mathbf{r}_0^N,\mathbf{p}_0^N] - c)\exp[-\beta H(\mathbf{r}_0^N,\mathbf{p}_0^N)] \tag{F.4}$$

from which the statistical mechanical probability distribution of c can be written:

$$\begin{aligned}
P(c) &= \frac{Q(c)}{Q} \\
&= \frac{\int \mathrm{d}\mathbf{r}_0^N \mathrm{d}\mathbf{p}_0^N \delta(c[x(t);\mathbf{r}_0^N,\mathbf{p}_0^N] - c)\exp[-\beta H(\mathbf{r}_0^N,\mathbf{p}_0^N)]}{Q}.
\end{aligned} \tag{F.5}$$

By comparing Equations F.3 and F.5 it is concluded that $\tilde{P}(c) = P(c)$, meaning that the probability distribution functions obtained from path sampling are true statistical mechanical ensemble distributions. An analogous relationship holds for the free energy as a function of $Q_6[x(t)]$. However, it should be noted that $\tilde{P}(Q_6[x(t)]) \neq P(Q_6(\mathbf{r}_0^N))$, as $Q_6[x(t)]$ is a path averaged quantity whereas $Q_6(\mathbf{r}_0^N)$

[1] If $c[x(t)]$ is measured from deterministic trajectories, the product in Equation F.1 is equal to unity and does not appear in Equation F.3

is the value of Q_6 for a given molecular configuration.[2] However, for the short path lengths studied in this work, where the system cannot change state on the timescale of the observation length, $\tilde{P}(Q_6[x(t)]) \approx P(Q_6(\mathbf{r}_0^N))$. The validity of this approximation is supported by the similarity of the barrier height in $F(Q_6[x(t)])$ at 260 K obtained in this work with the barrier height obtained at 273 K in Reference [145].

F.3 Definition of the q_{12} Order Parameter

In this section, the definition of the q_{12} local bond order parameter is outlined. The original definition is given in Reference [158].

The 25-dimensional local bond order vector $\tilde{\mathbf{q}}_{12}(i)$ around molecule i is defined as:

$$\tilde{q}_{12m}(i) = \frac{1}{N_i^{\mathrm{b}}} \sum_{j=1}^{N_i^{\mathrm{b}}} Y_{lm}(\hat{\mathbf{r}}_{ij}), \tag{F.6}$$

where N_i^{b} is the number of water molecules within 4.8 Å of molecule i.[3] A spatial coarse graining step following Lechner and Dellago [243] is then introduced:

$$q_{12m}(i) = \frac{1}{N_i^{\mathrm{b}} + 1} \sum_{j=0}^{N_i^{\mathrm{b}}} \tilde{q}_{12m}(j) \tag{F.7}$$

where the sum runs not only over the neighbours of i but also i itself. One can then calculate $d_{12}(i,j) = \hat{\mathbf{q}}_{12}(i) \cdot \hat{\mathbf{q}}_{12}(j)$, from which

$$n_{\mathrm{c}}(i) = \sum_{j=1}^{N_i^{\mathrm{b}}} \Gamma(d_{12}(i,j)) \tag{F.8}$$

is computed, where

$$\Gamma(x) = \begin{cases} 1, & \text{if } x > 0.75; \\ 0, & \text{otherwise.} \end{cases} \tag{F.9}$$

Finally, molecule i is defined as crystalline if $n_c(i) \geq 12$. This order parameter should be sensitive not only to molecules in a cubic or hexagonal ice-like environ-

[2] The relationship $\tilde{P}(Q_6[x(t)]) = P(Q_6[x(t)])$ still holds.

[3] In the original definition [158], N_i^{b} corresponded to the 16 nearest neighbours of molecule i. Based on the radial distribution function, the 4.8 Å cutoff corresponds to a distance where there are 16 nearest neighbours on average in the liquid.

ment, but also molecules belonging to the clathrate hydrate crystal structures, and ice 0. Further analysis could be performed if one wanted to distinguish between these crystal structures [158].

F.4 TPS Simulations with 1728 mW Molecules

All of the results presented in Chapter 6 were obtained from simulations of 216 mW molecules. In order to test the sensitivity of the results to finite size effects, TPS simulations like those performed in Section 6.4 were performed on a simulation cell with 1728 mW molecules at both 220 K and 260 K. Due to the added cost of these simulations and the scaling behaviour of the Q_6 order parameter, different settings to those reported in Section 6.4 were used. These are outlined below.

F.4.1 Initial Trajectories

The results presented in this section are from trajectories derived from two initial trajectories. The first initial trajectory was obtained by taking a trajectory from one of the TPS simulations at 260 K and replicating it in all three cartesian dimensions. The velocity components were rescaled by $\sqrt{\frac{220}{260}}$ in order to generate the initial trajectory at 220 K.[4] This trajectory will be referred to as the 'replicated-trajectory'. The second initial trajectory was obtained by melting a 1728 mW cell of ice at 370 K and rescaling the velocity components by $\sqrt{\frac{T}{370}}$ where $T = 220$ or 260. This trajectory will be referred to as the 'reverse-melt trajectory'. By using these two completely independent initial trajectories one is able to gauge the sensitivity of the results.

F.4.2 Settings

It was found that a 200 ps trajectory in a system of 1728 mW molecules is not long enough to connect the liquid and crystal basins such that the transition path ensemble could be effectively sampled. It is also significantly more expensive computationally. The main findings from the results with 216 mW molecules, however, were concerned with the early stages of nucleation i.e. whether or not the mobility changed before structuring and whether nucleation occurred in mobile

[4]The reason for not replicating a 220 K from the 216 mW simulations is that it proved difficult to obtain a trajectory starting in the liquid basin this way. This is due to the scaling behaviour of the Q_6 order parameter.

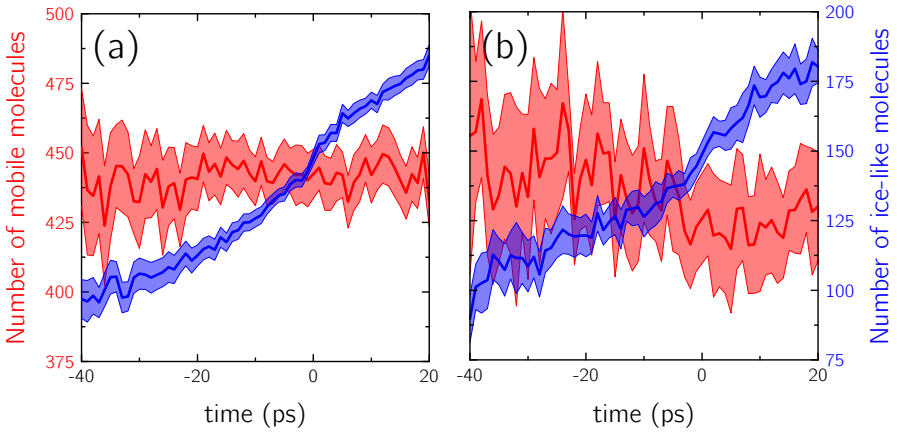

Figure F.1: Time evolution of the number of mobile and ice-like particles at 220 K in a 1728 mW system. Panel (a) shows results using the replicated-trajectory as an initial trajectory and panel (b) shows results using the reverse-melt trajectory as an initial trajectory. The results support the conclusions drawn from the 216 mW simulations: that the mobility does not change significantly during the initial stages of nucleation at this temperature. A value of $N_{\text{threshold}} = 120$ was used in this analysis.

or immobile domains. To this end, a trajectory was reactive if $Q_6[x(0)] < 0.017$ and $Q_6[x(\mathcal{T})] > 0.05$ (one should also bear in mind that the average value of Q_6 in the liquid goes as $1/N^2$ [145]), enabling shorter 75 ps trajectories to be used. Shooting and shifting moves were attempted in a ratio of 1:1, and the maximum shifting length was 12 ps. All other settings were the same as those reported in Section 6.4.

F.4.3 Results

In Figures F.1 and F.2 the time evolution of the number of mobile and ice-like molecules (determined by the local q_3 order parameter) at 220 K and 260 K are presented, respectively. The results are consistent with those obtained using 216 mW molecules. In Figure F.3, results from an analysis using the q_{12} order parameter at both temperatures are presented. Again, the results are consistent with those obtained with 216 mW molecules.

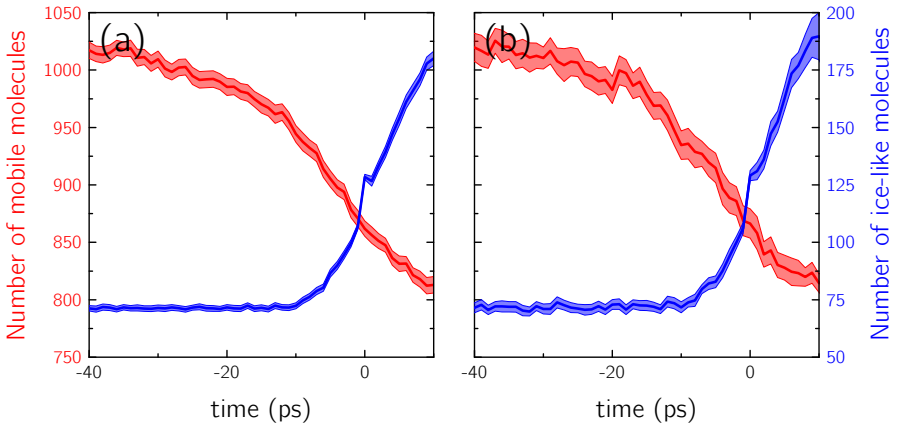

Figure F.2: Time evolution of the number of mobile and ice-like particles at 260 K in a 1728 mW system. Panel (a) shows results using the replicated-trajectory as an initial trajectory and panel (b) shows results using the reverse-melt trajectory as an initial trajectory. As observed for the 216 mW system, a clear drop in mobility is seen prior to ice formation. A value of $N_{\text{threshold}} = 160$ was used in this analysis.

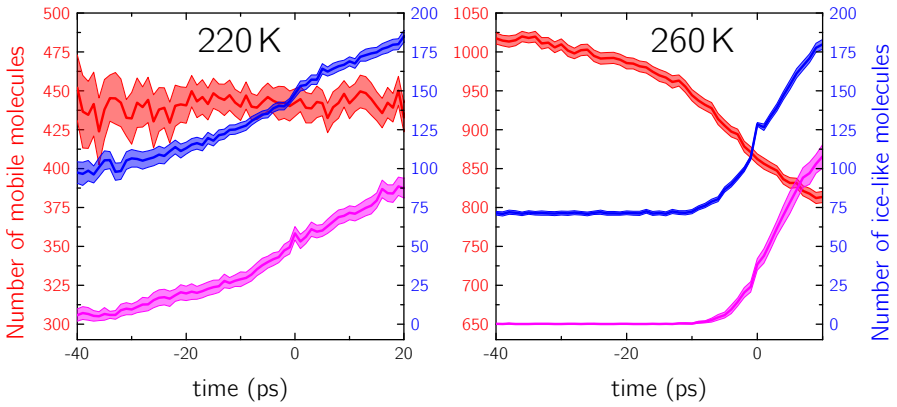

Figure F.3: Time evolution of other crystal structures in a 1728 mW system. The red and blue data are the same as in Figures F.1 and F.2 for 220 K and 260 K, respectively. The magenta line shows the number of crystal-like molecules using the q_{12} order parameter. Consistent with the results for the 216 mW system, there is no indication that ice 0 plays an important role in homogeneous nucleation.

Appendix G

Supplementary Material for Chapter 7

G.1 Overview

This appendix provides supplementary material to Chapter 7. Details regarding the force field parameters used are presented, as well results from other point charge force fields. The effect of long range corrections to the Lennard-Jones potential are discussed, as are the details of the fitting procedure used to determine the lattice constants and cohesive energies. Results for bulk ice I_h are also presented in more detail, and a derivation of the expression for $\Delta E_{\mathrm{diss}}^{\mathrm{sI}\rightarrow\mathrm{ice}}$ is provided (see Equation 7.20).

G.2 Force Field Parameters

To describe the water-water interactions, the TIP4P-2005 [91] and TIP4P-ICE [114] force fields (FFs) have been used. These FFs consist of a Lennard-Jones (LJ) potential centred on the oxygen atom of the water molecule and point charges located on the hydrogen atoms and a massless 'M' site along the H–H bisector. The parameters describing these parameters are given in Table G.1. In all of the calculations, the geometry of the water molecule has been constrained using the SETTLES algorithm [101].

For compatibility with the TIP4P-2005 and TIP4P-ICE potentials for water,

Potential	q_{M}	q_{H}	d_{OH}	d_{OM}	$\angle_{\mathrm{HOH}}$	ϵ_{OO}	σ_{OO}
TIP4P-2005	−1.1128	+0.5564	0.9572	0.15460	104.52°	0.7749	3.1589
TIP4P-ICE	−1.1794	+0.5897	0.9572	0.15139	104.52°	0.8822	3.1668

Table G.1: Parameters for the water FFs. The charges q_{M} and q_{H} are located on the massless M site and hydrogen atoms, respectively (given in atomic units). The distances d_{OH} and d_{OM} are the O–H and O–M bond lengths respectively (in Ångstrom) and $\angle_{\mathrm{HOH}}$ is the internal angle of the water molecule. The parameters ϵ_{OO} and σ_{OO} describe the strength and range of the LJ potential that is centred on the oxygen atom, and are given in kJ/mol and Ångstrom, respectively.

point charge and LJ potentials to describe the methane have also been chosen. More specifically, alongside with OPLS-AA [193] potential presented in Chapter 7, FFs that have already been reported in the literature have been tested. These include: a potential due to Tse, Klein and McDonald (TKM) [244]; the potential of Murad and Gubbins (MG) as specified in the article by Alavi *et al.* [245]; and the widely used Amber94 [246] and Charmm22 [247] FFs. Values for the point charges and LJ potentials for these FFs are given in Table G.2. Standard Lorentz-Berthelot mixing rules [97, 98] were used to calculate the LJ parameters between different atom types. Whereas the `SETTLES` algorithm provides a stable and efficient means to impose rigidity of the water molecules, to do so for methane would require the use of another algorithm such as `LINCS` [102] or `SHAKE` [248], which are unfortunately incompatible with the `L-BFGS` [212, 213] minimisation algorithm used in this work. Flexible potentials for the methane molecules have therefore been used. Intramolecular bond stretching between carbon and hydrogen atoms is described by a harmonic potential:

$$V_{\mathrm{b}}(r_{\mathrm{CH}}) = \frac{1}{2}k^{(\mathrm{b})}_{\mathrm{CH}}(r_{\mathrm{CH}} - b_{\mathrm{CH}})^2 \tag{G.1}$$

where r_{CH} is the distance between the carbon and hydrogen atoms, b_{CH} is the equilibrium bond length and $k^{(\mathrm{b})}_{\mathrm{CH}}$ is the force constant for the harmonic potential. The intramolecular H–C–H angle bend is also described by an harmonic potential:

$$V_{\mathrm{a}}(\theta_{\mathrm{HCH}}) = \frac{1}{2}k^{(\mathrm{a})}_{\mathrm{HCH}}(\theta_{\mathrm{HCH}} - \theta^0_{\mathrm{HCH}})^2 \tag{G.2}$$

where θ_{HCH} is the H–C–H angle, θ^0_{HCH} is the equilibrium value of θ_{HCH} and $k^{(\mathrm{a})}_{\mathrm{HCH}}$ is

Potential	q_C	q_H	ϵ_{CC}	ϵ_{HH}	σ_{CC}	σ_{HH}
OPLS-AA	−0.240	+0.060	0.276144	0.125520	3.5000	2.5000
TKM	−0.560	+0.140	1.365000	–	3.6400	–
MG	−0.572	+0.143	0.425660	0.071760	3.3500	2.6100
AMBER94	−0.464	+0.116	0.457730	0.0656888	3.3997	2.6495
CHARMM22	−0.360	+0.090	0.334720	0.0920480	3.6705	2.3520

Table G.2: Point charge and LJ parameters used for the methane FFs. The charges q_C and q_H are located on the carbon and hydrogen atoms, respectively (atomic units). With the exception of TKM, both carbon and hydrogen atoms act as LJ centres. The units of $\epsilon_{CC/HH}$ and $\sigma_{CC/HH}$ are kJ/mol and Ångstrom respectively.

Potential	$k_{CH}^{(b)}$	b_{CH}	$k_{HCH}^{(a)}$	θ_{HCH}^{0}
OPLS-AA	2845.120	1.0900	276.144	109.471°
TKM	2845.120	1.0940	276.144	109.471°
MG	2845.120	1.1000	276.144	109.471°
AMBER94	2845.120	1.0900	292.880	109.471°
CHARMM22	2694.496	1.1110	297.064	109.471°

Table G.3: Parameters used for the intramolecular bond stretching and angle bending terms of methane. The units of $k_{CH}^{(b)}$ and $k_{HCH}^{(a)}$ are kJ/mol Å^{-2} and kJ/mol rad^{-2} respectively. The form of the bond stretch and angle bend potentials are given in Equations G.1 and G.2.

the force constant. The values used for these parameters are given in Table G.3. For the TKM and MG potentials, where force constants were unavailable, the values from OPLS-AA have been used. Furthermore, structural parameters for the MG potential were taken from Reference [249]. In the next section, results from each of these methane potentials used in conjunction with the TIP4P-2005 and TIP4P-ICE water models are presented.

G.3 Results From Other Force Fields

In Table G.4 results from each of the FFs discussed above are presented. Although there is some variation in the values of a_0 and $\Delta E_{coh}^{sI}(a_0)$ obtained, all the FFs overbind the hydrate crystal but give reasonable structural parameters. From the findings in Chapter 7, this is to be expected as it is the water-water interactions that dominate the cohesive energy and structural properties of the hydrate. All the results for ΔE_{CH_4} and $\Delta E_{diss}^{sI\rightarrow ice}$ are also similar and in reasonable

Method	$\Delta E_{\mathrm{coh}}^{\mathrm{sI}}(a_0)$	a_0	$\Delta E_{\mathrm{CH_4}}$	$\Delta E_{\mathrm{coh}}^{\mathrm{empty}}$	$\Delta E_{\mathrm{diss}}^{\mathrm{sI}\rightarrow\mathrm{ice}}$
OPLS-AA/TIP4P-2005	-685	11.726	-248	-642	$+184$
TKM/TIP4P-2005	-686	11.726	-255	-642	$+190$
MG/TIP4P-2005	-683	11.730	-234	-642	$+169$
AMBER94/TIP4P-2005	-683	11.732	-239	-642	$+175$
CHARMM22/TIP4P-2005	-683	11.728	-236	-642	$+172$
OPLS-AA/TIP4P-ICE	-747	11.792	-261	-702	$+190$
TKM/TIP4P-ICE	-748	11.792	-268	-702	$+196$
MG/TIP4P-ICE	-744	11.795	-246	-702	$+174$
AMBER94/TIP4P-ICE	-745	11.797	-252	-702	$+180$
CHARMM22/TIP4P-ICE	-745	11.795	-248	-702	$+177$
DMC	-632 ± 1	11.83 ± 0.02	-241 ± 15	-590 ± 2	$+155 \pm 34$

Table G.4: Results for bulk sI methane hydrate obtained with different methane and water FF combinations. Although there is some sensitivity to the methane potential used, all of the results are broadly similar. The units are: $\Delta E_{\mathrm{coh}}^{\mathrm{sI}}(a_0)$ and $\Delta E_{\mathrm{coh}}^{\mathrm{empty}}$ in meV/H_2O; $\Delta E_{\mathrm{CH_4}}$ and $\Delta E_{\mathrm{diss}}^{\mathrm{sI}\rightarrow\mathrm{ice}}$ in meV/CH_4; and a_0 in Å.

agreement with the DMC values. All of these results have been obtained using long range corrections due to truncation of the LJ functions, but it will be seen in the following section that this only has a minor effect.

G.4 Effects of Tail Corrections

When truncating the LJ potential, it is possible to apply corrections to the potential energy due to neglect of interactions beyond the cutoff distance (for an outline of how this is implemented in `GROMACS 4.5.5` see Reference [250] and references therein). In Figure G.1 the results obtained with OPLS-AA/TIP4P-ICE both with and without tail correction applied are compared. When the tail correction is not applied, $\Delta E_{\mathrm{coh}}^{\mathrm{sI}}(a_0) = -742\,\mathrm{meV/H_2O}$, only 5 meV/$H_2O$ higher than the value reported in Table G.4 with the tail correction. The lattice constant is also very similar, with $a_0 = 11.797\,\text{Å}$, which is within 0.05% of the value reported in Table G.4. Furthermore, without tail corrections, $\Delta E_{\mathrm{CH_4}} = -250\,\mathrm{meV/CH_4}$, $\Delta E_{\mathrm{coh}}^{\mathrm{empty}} = -699\,\mathrm{meV/H_2O}$ and $\Delta E_{\mathrm{diss}}^{\mathrm{sI}\rightarrow\mathrm{ice}} = +181\,\mathrm{meV/CH_4}$ ($\Delta E_{\mathrm{coh}}^{\mathrm{ice}} = -711\,\mathrm{meV/H_2O}$ for TIP4P-ICE without tail corrections). It is concluded that the findings presented in Chapter 7 are therefore not significantly dependent on whether or not tail corrections are applied (and also likely the value

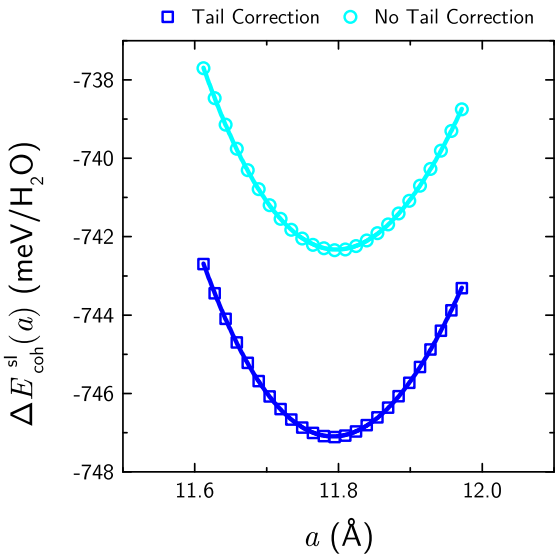

Figure G.1: Effects of tail corrections on the cohesive energy. This calculation has been performed with OPLS-AA/TIP4P-ICE. As expected, applying the tail correction makes the cohesive energy more exothermic. The effect is only small, however, with $\Delta E_{\mathrm{coh}}^{\mathrm{sI}}(a_0) = -742\,\mathrm{meV/H_2O}$, only $5\,\mathrm{meV/H_2O}$ higher than the value reported in Table G.4 with the tail correction. The equilibrium lattice constant is $a_0 = 11.797\,\text{Å}$ when the tail correction is not applied.

of the cutoff, provided a reasonable value is chosen).

G.5 Details of the Fitting Procedure

In order to determine $\Delta E_{\mathrm{coh}}^{\mathrm{sI}}(a_0)$ and $\Delta E_{\mathrm{coh}}^{\mathrm{ice}}$ the cohesive energy has been calculated for different volumes of the hydrate and ice crystals, and the resulting data points fitted to Murnaghan's equation of state [216, 217]:

$$\Delta E_{\mathrm{coh}}(V) = \Delta E_{\mathrm{coh}}(V_0) + \frac{B_0 V}{B_0'}\left(\frac{(V_0/V)^{B_0'}}{B_0' - 1} + 1\right) - \frac{V_0 B_0}{B_0' - 1} \qquad \text{(G.3)}$$

where V is the volume of the crystal, V_0 is the equilibrium volume of the crystal, B_0 is the bulk modulus and B_0' its pressure derivative. In Equation G.3, $\Delta E_{\mathrm{coh}}(V_0)$, V_0, B_0 and B_0' are four freely varying parameters that one can use to fit $E_{\mathrm{coh}}(V)$. In the case of methane hydrate, which has a cubic unit cell, V and V_0 are replaced by a^3 and a_0^3 respectively. To perform the fitting procedure, version 8.6 of the `OriginPro` software package from OriginLab, Northampton, MA was used. `OriginPro` uses the Levenberg-Marquardt algorithm to adjust the parameter values in an iterative procedure. It also reports standard errors in the fitted parameters by calculating the square root of the diagonal elements of the

variance-covariance matrix for the parameters. For the DMC calculations, each data point in Fig 7.1 (a) has been weighted during the fitting procedure according to the size of the 1 standard deviation error bars shown. The standard errors in fitted parameters $\Delta E_{\mathrm{coh}}^{\mathrm{sI}}(a_0)$ and a_0 reported by `OriginPro` are the errors reported in Table 7.1.

G.6 Results for Bulk Hexagonal Ice

In order to compute $\Delta E_{\mathrm{diss}}^{\mathrm{sI}\rightarrow\mathrm{ice}}$ it is necessary to compute $\Delta E_{\mathrm{coh}}^{\mathrm{ice}}$, the cohesive energy of ice $\mathrm{I_h}$. To do this, the procedure of References [175] and [176] has been followed, using the same 12 molecule unit cell of ice $\mathrm{I_h}$. With this unit cell, $\Delta E_{\mathrm{coh}}^{\mathrm{ice}}$ was then calculated at a number of different volumes of the crystal:

$$\Delta E_{\mathrm{coh}}^{\mathrm{ice}}(V) = \frac{\Delta E_{\mathrm{coh}}^{\mathrm{ice}}(V_0) - 12E_{\mathrm{H_2O}}}{12} \tag{G.4}$$

where $E_{\mathrm{H_2O}}$ is the energy of the isolated water molecule, V is the volume of the crystal and $\Delta E_{\mathrm{coh}}^{\mathrm{ice}}(V_0) \equiv \Delta E_{\mathrm{coh}}^{\mathrm{ice}}$ is the cohesive energy of the crystal at the equilibrium volume of the crystal V_0. For each volume of the crystal, the ratio of the ice $\mathrm{I_h}$ lattice constants (c/a) was maintained at the experimental value; this has previously been shown to be an excellent approximation for a variety of different xc functionals [44, 176]. Again these data from the DFT and FF calculations are fitted to Murnaghan's equation of state (Equation G.3). These results are presented in Table G.5, along with results from Reference [176] for comparison. Agreement with Reference [176] is fair, although $\Delta E_{\mathrm{coh}}^{\mathrm{ice}}$ is consistently slightly more exothermic and V_0 is slightly underestimated with the settings used here. This is most likely due to a different treatment of the core electrons between the two sets of calculations. Given the large differences from the DMC data, the strong correlation between the sI hydrate results and the ice $\mathrm{I_h}$ results (Figure 7.4) and the fact that the same settings have been used for all of the DFT calculations throughout the article, the general conclusions should be unaffected by these relatively small differences.

Method	$\Delta E_{\text{coh}}^{\text{ice}}$	V_0
This work:		
LDA	-1136	25.23
PBE	-657	30.36
PBE-D2	-758	29.04
PBE-vdW$^{\text{TS}}$	-737	29.25
revPBE-vdW	-583	33.87
optPBE-vdW	-696	31.20
optB88-vdW	-725	29.79
optB86b-vdW	-733	29.38
TIP4P-2005	-653	31.41
TIP4P-ICE	-714	31.95
From Ref. [176]:		
PBE	-636	30.79
PBE-vdW$^{\text{TS}}$	-714	29.67
revPBE-vdW	-559	34.38
optPBE-vdW	-668	31.63
DMC	-605 ± 5	31.69 ± 0.01

Table G.5: Cohesive energies and equilibrium volumes of ice I_h. DMC values are taken from Reference [175]. The results are in fair agreement with those reported in Reference [176]. The units of $\Delta E_{\text{coh}}^{\text{ice}}$ and V_0 are meV/H_2O and Å^3/H_2O, respectively.

G.7 Derivation of Expression for $\Delta E_{\text{diss}}^{\text{sI}\to\text{ice}}$

In this last section, a derivation of Equation 7.20 used to calculate $\Delta E_{\text{diss}}^{\text{sI}\to\text{ice}}$ is presented. Rather than calculating $E_{\text{diss}}^{\text{sI}\to\text{ice}}$ explicitly from Equation 7.19, instead consider the following sequences of processes:

$$5.75H_2O \cdot CH_{4(\text{sI})} \xrightarrow{-\Delta E_{CH_4}} 5.75H_2O_{(\text{sI})} + CH_{4(\text{gas})} \tag{G.5}$$

$$5.75H_2O_{(\text{sI})} + CH_{4(\text{gas})} \xrightarrow{-5.75\Delta E_{\text{empty}}} 5.75H_2O_{(\text{gas})} + CH_{4(\text{gas})} \tag{G.6}$$

$$5.75H_2O_{(\text{gas})} + CH_{4(\text{gas})} \xrightarrow{5.75\Delta E_{\text{ice}}} 5.75H_2O_{(\text{ice})} + CH_{4(\text{gas})} \tag{G.7}$$

It is then easy to see that:

$$\Delta E_{\text{diss}}^{\text{sI}\to\text{ice}} = 5.75\Delta E_{\text{ice}} - 5.75\Delta E_{\text{empty}} - \Delta E_{CH_4} \tag{G.8}$$

which is Equation 7.20. The main advantage of this expression for $\Delta E_{\mathrm{diss}}^{\mathrm{sI}\rightarrow\mathrm{ice}}$ is that one can use literature values [175] for the DMC number rather than performing another set of computationally expensive calculations. This form of the equation also allows one to directly compare the performance of the DFT/FF methods for both the hydrate and ice $\mathrm{I_h}$.

Bibliography

[1] H. R. Pruppacher and J. D. Klett. *Microphysics of Clouds and Precipitation - Second Revised and Enlarged Edition with an Introduction to Cloud Chemistry and Cloud Electricity.* Kluwer Academic Publishers, Dordrecht, The Netherlands, 1997.

[2] T. F. Stocker, D. Qin, G. K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley. *Climate Change 2013: The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

[3] E. Dendy Sloan and Carolyn A. Koh. *Clathrate Hydrates of Natural Gases, Third Edition.* CRC Press, Boca Raton, Florida, USA, 2008.

[4] E. D. Sloan. Fundamental principles and applications of natural gas hydrates. *Nature*, 426:353–359, 2003.

[5] Thorsten Bartels-Rausch. Chemistry: Ten things we need to know about ice and snow. *Nature*, 494:27–29, 2013.

[6] Thorsten Bartels-Rausch, Vance Bergeron, Julyan H. E. Cartwright, Rafael Escribano, John L. Finney, Hinrich Grothe, Pedro J. Gutierrez, Jari Haapala, Werner F. Kuhs, Jan B. C. Pettersson, Stephen D. Price, C. Ignacio Sainz-Diaz, Debbie J. Stokes, Giovanni Strazzulla, Erik S. Thomson, Hauke Trinks, and Nevin Uras-Aytemiz. Ice structures, patterns, and processes: A view across the icefields. *Rev. Mod. Phys.*, 84:885–944, 2012.

[7] B. J. Murray, D. O'Sullivan, J. D. Atkinson, and M. E. Webb. Ice nucleation by particles immersed in supercooled cloud droplets. *Chem. Soc. Rev.*, 41:6519–6554, 2012.

[8] K. Diehl and S.K. Mitra. A laboratory study of the effects of a kerosene-burner exhaust on ice nucleation and the evaporation rate of ice crystals. *Atmos. Environ.*, 32:3145–3151, 1998.

[9] Paul J. DeMott. An exploratory study of ice nucleation by soot aerosols. *J. Appl. Meteor.*, 29:1072 – 1079, 1990.

[10] P. J. DeMott, Y. Chen, S. M. Kreidenweis, D. C. Rogers, and D. E. Sherman. Ice formation by black carbon particles. *Geophys. Res. Lett.*, 26:2429–2432, 1999.

[11] K Diehl, S Matthias-Maser, R Jaenicke, and S. K. Mitra. The ice nucleating ability of pollen: Part II. Laboratory studies in immersion and contact freezing modes. *Atmos. Res,*, 61:125–133, 2002.

[12] Nadine von Blohn, Subir K. Mitra, Karoline Diehl, and Stephan Borrmann. The ice nucleating ability of pollen: Part III: New laboratory studies in immersion and contact freezing modes including more pollen types. *Atmos. Res.*, 78:182–189, 2005.

[13] B. G. Pummer, H. Bauer, J. Bernardi, S. Bleicher, and H. Grothe. Suspendable macromolecules are responsible for ice nucleation activity of birch and conifer pollen. *Atmos. Chem. Phys.*, 12:2541–2550, 2012.

[14] Leroy R Maki, Elizabeth L Galyan, Mei-Mon Chang-Chien, and Daniel R Caldwell. Ice nucleation induced by Pseudomonas syringae. *Appl. Microbiol.*, 28:456–459, 1974.

[15] R. P. Sear. The non-classical nucleation of crystals: microscopic mechanisms and applications to molecular crystals, ice and calcium carbonate. *Int. Mater. Rev.*, 57:328–356, 2012.

[16] J. A. Sellberg, C. Huang, T. A. McQueen, N. D. Loh, H. Laksmono, D. Schlesinger, R. G. Sierra, D. Nordlund, C. Y. Hampton, D. Starodub, D. P. DePonte, M. Beye, C. Chen, A. V. Martin, A. Barty, K. T. Wikfeldt, T. M. Weiss, C. Caronna, J. Feldkamp, L. B. Skinner, M. M. Seibert, M. Messerschmidt, G. J. Williams, S. Boutet, L. G. M. Pettersson, M. J. Bogan, and A. Nilsson. Ultrafast x-ray probing of water structure below the homogeneous ice nucleation temperature. *Nature*, 510:381–384, 2014.

[17] James D Atkinson, Benjamin J Murray, Matthew T Woodhouse, Thomas F Whale, Kelly J Baustian, Kenneth S Carslaw, Steven Dobbie, Daniel O'Sullivan, and Tamsin L Malkin. The importance of feldspar for ice nucleation by mineral dust in mixed-phase clouds. *Nature*, 498:355–358, 2013.

[18] Javier Carrasco, Andrew Hodgson, and Angelos Michaelides. A molecular perspective of water at metal interfaces. *Nature Mater.*, 11:667–674, 2012.

[19] Stephen J. Cox, Shawn M. Kathmann, John A. Purton, Michael J. Gillan, and Angelos Michaelides. Non-hexagonal ice at hexagonal surfaces: the role of lattice mismatch. *Phys. Chem. Chem. Phys.*, 14:7944–7949, 2012.

[20] Stephen J. Cox, Zamaan Raza, Shawn M. Kathmann, Ben Slater, and Angelos Michaelides. The microscopic features of heterogeneous ice nucleation may affect the macroscopic morphology of atmospheric ice crystals. *Faraday Discuss.*, 167:389–403, 2013.

[21] Stephen J. Cox, Michael D. Towler, Dario Alfè, and Angelos Michaelides. Benchmarking the performance of density functional theory and point charge force fields in their description of sI methane hydrate against diffusion Monte Carlo. *J. Chem. Phys.*, 140:174703, 2014.

[22] J. K. Nørskov, T. Bligaard, J. Rossmeisl, and C. H. Christensen. Towards the computational design of solid catalysts. *Nature Chem.*, 1:37–46, 2009.

[23] Martin P. Andersson, Thomas Bligaard, Arkady Kustov, Kasper E. Larsen, Jeffrey Greeley, Tue Johannessen, Claus H. Christensen, and Jens K.

Nørskov. Toward computational screening in heterogeneous catalysis: Pareto-optimal methanation catalysts. *J. Catal.*, 239:501 – 506, 2006.

[24] David Chandler. *Introduction to Modern Statistical Mechanics.* Oxford University Press, 198 Madison Avenue, New York, New York, U.S.A., 1987.

[25] Daan Frenkel and Berend Smit. *Understanding Molecular Simulation, From Algorithms to Applications.* Academic Press, San Diego, USA, second edition, 2002.

[26] Donald Allan McQuarrie. *Statistical Mechanics.* University Science Books, 55D Gate Five Road, Sausalito, CA 94965, 2000.

[27] Mark E. Tuckerman. *Statistical Mechanics: Theory and Molecular Simulation.* Oxford University Press, New York, USA, 2010.

[28] Robert H Swendsen. Statistical mechanics of colloids and Boltzmann's definition of the entropy. *Am J. Phys.*, 74:187–190, 2006.

[29] Robert H Swendsen. Gibbs' paradox and the definition of entropy. *Entropy*, 10:15–18, 2008.

[30] Daan Frenkel. Why colloidal systems can be described by statistical mechanics: some not very original comments on the Gibbs paradox. *Mol. Phys.*, 112:2325–2329, 2014.

[31] S Auer and D Frenkel. Prediction of absolute crystal-nucleation rate in hard-sphere colloids. *Nature*, 409:1020–1023, 2001.

[32] Shawn M. Kathmann. Understanding the chemical physics of nucleation. *Theor. Chem. Acc.*, 116:169–182, 2006.

[33] David Turnball and Bernard Vonnegut. Nucleation catalysis. *Ind. Eng. Chem.*, 44:1292–1298, 1952.

[34] Bilal Zuberi, Allan K. Bertram, Thomas Koop, Luisa T. Molina, and Mario J. Molina. Heterogeneous freezing of aqueous particles induced by crystallized $(NH_4)_2SO_4$, ice, and letovicite. *J. Phys. Chem. A*, 105(26):6458–6464, 2001.

[35] R. Popovitz-Biro, J. L. Wang, J. Majewski, E. Shavit, L. Leiserowitz, and M. Lahav. Induced freezing of supercooled water into ice by self-assembled crystalline monolayers of amphiphilic alcohols at the air-water interface. *J. Am. Chem. Soc.*, 116:1179–1191, 1994.

[36] Benjamin J. Murray, Theodore W. Wilson, Steven Dobbie, Zhiqiang Cui, Sardar M. R. K. Al-Jumur, Ottmar Möhler, Martin Schnaiter, Robert Wagner, Stefan Benz, Monika Niemand, Harald Saathoff, Volker Ebert, Steven Wagner, and Bernd Kaercher. Heterogeneous nucleation of ice particles on glassy aerosols under cirrus conditions. *Nature Geosci.*, 3:233–237, 2010.

[37] D. A. Knopf, B. Wang, A. Laskin, R. C. Moffet, and M. K. Gilles. Heterogeneous nucleation of ice on anthropogenic organic particles collected in mexico city. *Geophy. Res. Lett.*, 37:L11803, 2010.

[38] Stefano Curtarolo, Gus LW Hart, Marco Buongiorno Nardelli, Natalio Mingo, Stefano Sanvito, and Ohad Levy. The high-throughput highway to computational materials design. *Nature Mater.*, 12:191–201, 2013.

[39] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford: Clarendon Press, 1989.

[40] B. Vonnegut. The nucleation of ice formation by silver iodide. *J. Appl. Phys.*, 18:593–595, 1947.

[41] D. L. Doering and T. E. Madey. The adsorption of water on clean and oxygen-dosed Ru(011). *Surf. Sci.*, 123:305–337, 1982.

[42] Xiao Liang Hu and Angelos Michaelides. Ice formation on kaolinite: Lattice match or amphoterism? *Surf. Sci.*, 601:5378–5381, 2007.

[43] T. Croteau, A. K. Bertram, and G. N. Patey. Adsorption and structure of water on kaolinite surfaces: Possible insight into ice nucleation from grand canonical Monte Carlo calculations. *J. Phys. Chem. A*, 112:10708–10712, 2008.

[44] Peter J. Feibelman. Lattice match in density functional calculations: ice I_h vs. β-AgI. *Phys. Chem. Chem. Phys.*, 10:4688–4691, 2008.

[45] A Michaelides, A Alavi, and D A King. Insight into H_2O-ice adsorption and dissociation on metal surfaces from first-principles simulations. *Phys. Rev. B*, 69:113404, 2004.

[46] S. Nie, Peter J. Feibelman, N. C. Bartelt, and K. Thürmer. Pentagons and heptagons in the first water layer on Pt(111). *Phys. Rev. Lett.*, 105:026102, 2010.

[47] Sebastian Standop, Alex Redinger, Markus Morgenstern, Thomas Michely, and Carsten Busse. Molecular structure of the H_2O wetting layer on Pt(111). *Phys. Rev. B*, 82:161412, 2010.

[48] Mark Gallagher, Ahmed Omer, George R. Darling, and Andrew Hodgson. Order and disorder in the wetting layer on Ru(0001). *Faraday Discuss.*, 141:231–249, 2009.

[49] Javier Carrasco, Angelos Michaelides, Matthew Forster, Sam Haq, Rasmita Raval, and Andrew Hodgson. A one-dimensional ice structure built from pentagons. *Nature Mater.*, 8:427, 2009.

[50] David R. Nutt and Anthony J. Stone. Ice nucleation on a model hexagonal surface. *Langmuir*, 20:8715–8720, 2004.

[51] H. Witek and V. Buch. Structure of ice multilayers on metals. *J. Chem. Phys.*, 110:3168–3175, 1999.

[52] T Lankau and I L Cooper. $(H_2O)_6$ on a virtual metal surface: Testing the surface ice rules. *J. Phys. Chem A*, 105:4084–4095, 2001.

[53] Henning Henschel, Tobias Kraemer, and Timm Lankau. $(H_2O)_{10}$ and $(H_2O)_{12}$ on a virtual metal surface: The growth of ice. *Langmuir*, 22:10942–10950, 2006.

[54] Fiona McBride, George R. Darling, Katariina Pussi, and Andrew Hodgson. Tailoring the structure of water at a metal surface: A structural analysis of

the water bilayer formed on an alloy template. *Phys. Rev. Lett.*, 106:226101, 2011.

[55] Matthew Forster, Rasmita Raval, Andrew Hodgson, Javier Carrasco, and Angelos Michaelides. c(2×2) water-hydroxyl layer on Cu(110): A wetting layer stabilized by Bjerrum defects. *Phys. Rev. Lett.*, 106:046103, 2011.

[56] Angelos Michaelides and Karina Morgenstern. Ice nanoclusters at hydrophobic metal surfaces. *Nature Mater.*, 6:597–601, 2007.

[57] David J Wales and Matthew P Hodges. Global minima of water clusters $(H_2O)_n$, $n \leq 21$, described by an empirical potential. *Chem. Phys. Lett.*, 286:65, 1998.

[58] Jiří Klimeš, David R. Bowler, and Angelos Michaelides. Chemical accuracy for the van der Waals density functional. *J. Phys.: Condens. Matter*, 22:022201, 2010.

[59] Jiří Klimeš, David R. Bowler, and Angelos Michaelides. Van der Waals density functionals applied to solids. *Phys. Rev. B*, 83:195131, 2011.

[60] Javier Carrasco, Jiří Klimeš, and Angelos Michaelides. The role of van der Waals forces in water adsorption on metals. *J. Chem. Phys.*, 138:024708, 2013.

[61] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926, 1983.

[62] E. Sanz, C. Vega, J. L. F. Abascal, and L. G. MacDowell. Phase diagram of water from computer simulation. *Phys. Rev. Lett.*, 92:255701, 2004.

[63] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *J. Phys. Chem.*, 91:6269–6271, 1987.

[64] John Purton. `DL_MONTE`: Monte Carlo simulation package, 2011.

[65] Berk Hess, Carsten Kutzner, David van der Spoel, and Erik Lindahl. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theor. Comput.*, 4:435, 2008.

[66] Xiao Liang Hu and Angelos Michaelides. Water on the hydroxylated (001) surface of kaolinite: From monomer adsorption to a flat 2D wetting layer. *Surf. Sci.*, 602:960–974, 2008.

[67] V Sadtchenko, G E Ewing, D R Nutt, and A J Stone. Instability of ice films. *Langmuir*, 18:4632–4636, 2002.

[68] M Matsumoto, S Saito, and I Ohmine. Molecular dynamics simulation of the ice nucleation and growth process leading to water freezing. *Nature*, 416:409–413, 2002.

[69] Luboš Vrbka and Pavel Jungwirth. Homogeneous freezing of water starts in the subsurface. *J. Phys. Chem. B*, 110:18126, 2006.

[70] Eva Pluhařová, Luboš Vrbka, and Pavel Jungwirth. Effect of surface pollution on homogeneous ice nucleation: A molecular dynamics study. *J. Phys. Chem. C*, 114:7831, 2010.

[71] Sigurd Bauerecker, Peter Ulbig, Victoria Buch, Luboš Vrbka, and Pavel Jungwirth. Monitoring ice nucleation in pure and salty water via high-speed imaging and computer simulations. *J. Phys. Chem. C*, 112:7631–7636, 2008.

[72] Emily B Moore and Valeria Molinero. Is it cubic? Ice crystallization from deeply supercooled water. *Phys. Chem. Chem. Phys.*, 13:20008–20016, 2011.

[73] Emily B Moore and Valeria Molinero. Ice crystallization in water's "no-man's land". *J. Chem. Phys.*, 132:244504, 2010.

[74] Aleks Reinhardt and Jonathan P. K. Doye. Free energy landscapes for homogeneous nucleation of ice for a monatomic water model. *J. Chem. Phys.*, 136:054501, 2012.

[75] R Radhakrishnan and B L Trout. Nucleation of hexagonal ice I_h in liquid water. *J. Am. Chem. Soc.*, 125:7743–7747, 2003.

[76] D. Quigley and P. M. Rodger. Metadynamics simulations of ice nucleation and growth. *J. Chem. Phys.*, 128:154518, 2008.

[77] J. Y. Yan and G. N. Patey. Heterogeneous ice nucleation induced by electric fields. *J. Phys. Chem. Lett.*, 2:2555–2559, 2011.

[78] J Y Yan and G N Patey. Molecular dynamics simulations of ice nucleation by electric fields. *J. Phys. Chem. A*, 116:7057–7064, 2012.

[79] Estefanía González Solveyra, Ezequiel de la Llave, Damián A Scherlis, and Valeria Molinero. Melting and crystallization of ice in partially filled nanopores. *J. Phys. Chem. B*, 115:14196–14204, 2011.

[80] Valeria Molinero and Emily B. Moore. Water modeled as an intermediate element between carbon and silicon. *J. Phys. Chem. B*, 113:4008–4016, 2009.

[81] A. Laio and M Parrinello. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA*, 99:12562, 2002.

[82] Alessandro Laio and Francesco L Gervasio. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys.*, 71:126601, 2008.

[83] G.M. Torrie and J.P. Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.*, 23:187–199, 1977.

[84] Peter A. Smithson. IPCC, 2001: climate change 2001: the scientific basis. contribution of Working Group 1 to the Third Assessment Report of the Intergovernmental Panel on Climate Change, edited by J T Houghton, Y. Ding, D J Griggs, M. Noguer, P J van der Linden, X. Dai, K. Maskell and C A Johnson (eds). *Int. J. Climatol.*, 22:1144–1144, 2002.

[85] Michael L. Eastwood, Sebastien Cremel, Clemens Gehrke, Eric Girard, and Allan K. Bertram. Ice nucleation on mineral dust particles: Onset condi-

tions, nucleation rates and contact angles. *J. Geophys. Res.*, 113:D22203, 2008.

[86] Frank Zimmermann, Stephan Weinbruch, Lothar Schuetz, Heiko Hofmann, Martin Ebert, Konrad Kandler, and Annette Worringen. Ice nucleation properties of the most abundant mineral dust phases. *J. Geophys. Res.*, 113:D23204, 2008.

[87] B. J. Murray, S. L. Broadley, T. W. Wilson, J. D. Atkinson, and R. H. Wills. Heterogeneous freezing of water droplets containing kaolinite particles. *Atmos. Chem. Phys.*, 11:4191–4207, 2011.

[88] Xiao Liang Hu and Angelos Michaelides. The kaolinite (001) polar basal plane. *Surf. Sci.*, 604:111–117, 2010.

[89] T. Croteau, A. K. Bertram, and G. N. Patey. Simulation of water adsorption on kaolinite under atmospheric conditions. *J. Phys. Chem. A*, 113:7826–7833, 2009.

[90] T. Croteau, A. K. Bertram, and G. N. Patey. Water adsorption on kaolinite surfaces containing trenches. *J. Phys. Chem. A*, 114:2171–2178, 2010.

[91] J. L. F. Abascal and C. Vega. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.*, 123:234505, 2005.

[92] R T Cygan, J J Liang, and A G Kalinichev. Molecular models of hydroxide, oxyhydroxide, and clay phases and the development of a general force field. *J. Phys. Chem. B*, 108:1255, 2004.

[93] James P. Larentzos, Jeffery A. Greathouse, and Randall T. Cygan. An *ab initio* and classical molecular dynamics investigation of the structural and vibrational properties of talc and pyrophyllite. *J. Phys. Chem. C*, 111:12752–12759, 2007.

[94] Jianwei Wang, Andrey G. Kalinichev, R. James Kirkpatrick, and Randall T. Cygan. Structure, energetics, and dynamics of water adsorbed on the mus-

covite (001) surface: A molecular dynamics simulation. *J. Phys. Chem. B*, 109:15893–15905, 2005.

[95] Dimitrios Argyris, David R. Cole, and Alberto Striolo. Ion-specific effects under confinement: The role of interfacial water. *ACS Nano*, 4:2035–2042, 2010.

[96] Benjamin Rotenberg, Amish J. Patel, and David Chandler. Molecular explanation for why talc surfaces can be both hydrophilic and hydrophobic. *J. Am. Chem. Soc.*, 133:20521–20527, 2011.

[97] H. A. Lorentz. Ueber die anwendung des satzes vom virial in der kinetischen theorie der gase. *Ann. Phys.*, 248:127–136, 1881.

[98] D. Berthelot. Sur le mélange des gaz. *C. R. Hebd. Séanc. Acad. Sci.*, 126:1703–1855, 1898.

[99] D. L. Bish. Rietveld refinement of the kaolinite structure at 1.5 K. *Clays Clay Miner.*, 41:738–744, 1993.

[100] I C Yeh and M L Berkowitz. Ewald summation for systems with slab geometry. *J. Chem. Phys.*, 111:3155–3162, 1999.

[101] Shuichi Miyamoto and Peter A. Kollman. `SETTLE`: An analytical version of the `SHAKE` and `RATTLE` algorithm for rigid water models. *J. Comput. Chem.*, 13:952–962, 1992.

[102] Berk Hess, Henk Bekker, Herman J. C. Berendsen, and Johannes G. E. M. Fraaije. `LINCS`: a linear constraint solver for molecular simulations. *J. Comput. Chem.*, 18:1463–1472, 1997.

[103] `http://pubs.rsc.org/en/content/articlelanding/2013/fd/c3fd00059a`.

[104] Tamsin L. Malkin, Benjamin J. Murray, Andrey V. Brukhno, Jamshed Anwar, and Christoph G. Salzmann. Structure of ice crystallized from supercooled water. *Proc. Natl. Acad. Sci. USA*, 2012.

[105] Werner F Kuhs, Christian Sippel, Andrzej Falenty, and Thomas C Hansen. Extent and relevance of stacking disorder in "ice I_c". *Proc. Natl. Acad. Sci. USA*, 109:21259–21264, 2012.

[106] Marcelo A. Carignano. Formation of stacking faults during ice growth on hexagonal and cubic substrates. *J. Phys. Chem. C*, 111:501–504, 2007.

[107] Ding Pan, Li-Min Liu, Ben Slater, Angelos Michaelides, and Enge Wang. Melting the ice: on the relation between melting temperature and size for nanoscale ice crystals. *ACS Nano*, 5:4562–4569, 2011.

[108] C. L. Bishop, D. Pan, L. M. Liu, G. A. Tribello, A. Michaelides, E. G. Wang, and B. Slater. On thin ice: Surface order and disorder during pre-melting. *Faraday Discuss.*, 141:277–292, 2009.

[109] H. Nada and Y. Furukawa. Anisotropic growth kinetics of ice crystals from water studied by molecular dynamics simulation. *J Cryst. Growth*, 169:587–597, 1996.

[110] Dmitri Rozmanov and Peter G. Kusalik. Anisotropy in the crystal growth of hexagonal ice I_h. *J. Chem. Phys.*, 137:094702, 2012.

[111] Emily B. Moore, Ezequiel de la Llave, Kai Welke, Damian A. Scherlis, and Valeria Molinero. Freezing, melting and structure of ice in a hydrophilic nanopore. *Phys. Chem. Chem. Phys.*, 12:4124–4134, 2010.

[112] M. Salmeron, H. Bluhm, M. Tatarkhanov, G. Ketteler, T. K. Shimizu, A. Mugarza, Xingyi Deng, T. Herranz, S. Yamamoto, and A. Nilsson. Water growth on metals and oxides: binding, dissociation and role of hydroxyl groups. *Faraday Discuss.*, 141:221–229, 2009.

[113] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, 1996.

[114] J. L. F. Abascal, E. Sanz, R. García Fernández, and C. Vega. A potential model for the study of ices and amorphous water: TIP4P/Ice. *J. Chem. Phys.*, 122:234511, 2005.

[115] Matthew P. Bailey and John Hallett. A comprehensive habit diagram for atmospheric ice crystals: Confirmation from the laboratory, AIRS II, and other field studies. *J. Atmos. Sci.*, 66:2888–2899, 2009.

[116] Ding Pan, Li-Min Liu, Gareth A. Tribello, Ben Slater, Angelos Michaelides, and Enge Wang. Surface energy and surface proton order of the ice I_h basal and prism surfaces. *J. Phys.: Condensed Matter*, 22:074209, 2010.

[117] Aleks Reinhardt, Jonathan P. K. Doye, Eva G. Noya, and Carlos Vega. Local order parameters for use in driving homogeneous ice nucleation with all-atom models of water. *J. Chem. Phys.*, 137:194504, 2012.

[118] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Ann. Rev. Phys. Chem.*, 53:291–318, 2002.

[119] Rosalind J Allen, Daan Frenkel, and Pieter Rein Ten Wolde. Simulating rare events in equilibrium or nonequilibrium stochastic systems. *J. Chem. Phys.*, 124:024102, 2006.

[120] Rosalind J. Allen, Chantal Valeriani, and Pieter Rein ten Wolde. Forward flux sampling for rare event simulations. *J. Phys.: Condens. Matter*, 21:463102, 2009.

[121] Emily B. Moore and Valeria Molinero. Structural transformation in supercooled water controls the crystallization rate of ice. *Nature*, 479:506–U226, 2011.

[122] Tianshu Li, Davide Donadio, Giovanna Russo, and Giulia Galli. Homogeneous ice nucleation from supercooled water. *Phys. Chem. Chem. Phys.*, 13:19807–19813, 2011.

[123] F. Stillinger and T. A. Weber. Computer simulation of local order in condensed phases of silicon. *Phys. Rev. B*, 31:5262, 1985.

[124] Azar Alizadeh, Masako Yamada, Ri Li, Wen Shang, Shourya Otta, Sheng Zhong, Liehui Ge, Ali Dhinojwala, Ken R. Conway, Vaibhav Bahadur,

A. Joseph Vinciquerra, Brian Stephens, and Margaret L. Blohm. Dynamics of ice nucleation on water repellent surfaces. *Langmuir*, 28:3180–3186, 2012.

[125] Kaiyong Li, Shun Xu, Wenxiong Shi, Min He, Huiling Li, Shuzhou Li, Xin Zhou, Jianjun Wang, and Yanlin Song. Investigating the effects of solid surfaces on ice nucleation. *Langmuir*, 28:10749–10754, 2012.

[126] Kaiyong Li, Shun Xu, Jing Chen, Qiaolan Zhang, Yifan Zhang, Dapeng Cui, Xin Zhou, Jianjun Wang, and Yanlin Song. Viscosity of interfacial water regulates ice nucleation. *Appl. Phys. Lett.*, 104:101605, 2014.

[127] S. Plimpton. Fast parallel algorithms for short-range molecular dynamics. *J. Comp. Phys.*, 117:1–19, 1995.

[128] S. R. Bahn and K. W. Jacobsen. An object-oriented scripting interface to a legacy electronic structure code. *Comput. Sci. Eng.*, 4:56–66, 2002.

[129] Laura Lupi, Arpa Hudait, and Valeria Molinero. Heterogeneous nucleation of ice on carbon surfaces. *J. Am. Chem. Soc.*, 136:3156–3164, 2014.

[130] Laura Lupi and Valeria Molinero. Does hydrophilicity of carbon particles improve their ice nucleation ability? *J. Phys. Chem. A*, 118:7330–7337, 2014.

[131] Tianjun Sun, Feng-Hsu Lin, Robert L. Campbell, John S. Allingham, and Peter L. Davies. An antifreeze protein folds with an interior network of more than 400 semi-clathrate waters. *Science*, 343:795–798, 2014.

[132] F Sicheri and DSC Yang. Ice-binding structure and mechanism of an antifreeze protein from winter flounder. *Nature*, 375:427–431, 1995.

[133] F. D. Sonnichsen, C. I. DeLuca, P. L. Davies, and B. D. Sykes. Refined solution structure of type III antifreeze protein: Hydrophobic groups may be involved in the energetics of the protein-ice interaction. *Structure*, 4:1325–1337, 1996.

[134] Y. C. Liou, A. Tocilj, P. L. Davies, and Z. C. Jia. Mimicry of ice structure by surface hydroxyls and water of a β-helix antifreeze protein. *Nature*, 406:322–324, 2000.

[135] S. P. Graether, M. J. Kuiper, S. M. Gagne, V. K. Walker, Z. C. Jia, B. D. Sykes, and P. L. Davies. β-helix structure and ice-binding properties of a hyperactive antifreeze protein from an insect. *Nature*, 406:325–328, 2000.

[136] Brad L. Pentelute, Zachary P. Gates, Valentina Tereshko, Jennifer L. Dashnau, Jane M. Vanderkooi, Anthony A. Kossiakoff, and Stephen B. H. Kent. X-ray structure of snow flea antifreeze protein determined by racemic crystallization of synthetic protein enantiomers. *J. Am. Chem. Soc.*, 130:9695–9701, 2008.

[137] J. Juraszek, G. Saladino, T. S. van Erp, and F. L. Gervasio. Efficient numerical reconstruction of protein folding kinetics with partial path sampling and pathlike variables. *Phys. Rev. Lett.*, 110:108106, 2013.

[138] Phillip L Geissler, Christoph Dellago, David Chandler, Jürg Hutter, and Michele Parrinello. Autoionization in liquid water. *Science*, 291:2121–2124, 2001.

[139] Sanford M Simon, Charles S Peskin, and George F Oster. What drives the translocation of proteins? *Proc. Natl. Acad. Sci. U.S.A.*, 89:3770–3774, 1992.

[140] Christoph Dellago, Peter G. Bolhuis, and David Chandler. Efficient transition path sampling: Application to Lennard-Jones cluster rearrangements. *J. Chem. Phys.*, 108:9236–9245, 1998.

[141] C Dellago, PG Bolhuis, and PL Geissler. Transition path sampling. In *Advances in Chemical Physics*, volume 123 of *Adv. Chem. Phys.*, pages 1–78. 2002.

[142] Peter G Bolhuis. Transition path sampling on diffusive barriers. *J. Phys.: Condens. Matter*, 15:S113, 2003.

[143] David Quigley and P Mark Rodger. A metadynamics-based approach to sampling crystallisation events. *Mol. Sim.*, 35:613–623, 2009.

[144] Aleks Reinhardt, Jonathan P. K. Doye, Eva G. Noya, and Carlos Vega. Local order parameters for use in driving homogeneous ice nucleation with all-atom models of water. *J. Chem. Phys.*, 137:194504, 2012.

[145] David T. Limmer and David Chandler. The putative liquid-liquid transition is a liquid-solid transition in atomistic models of water. *J. Chem. Phys.*, 135:134503, 2011.

[146] Lester O. Hedges, Robert L. Jack, Juan P. Garrahan, and David Chandler. Dynamic order-disorder in atomistic models of structural glass formers. *Science*, 323:1309–1313, 2009.

[147] Aaron S. Keys, Lester O. Hedges, Juan P. Garrahan, Sharon C. Glotzer, and David Chandler. Excitations are localized and relaxation is hierarchical in glass-forming liquids. *Phys. Rev. X*, 1:021013, 2011.

[148] David T Limmer and David Chandler. Theory of amorphous ices. *Proc. Natl. Acad. Sci. U.S.A*, 111:9413–9418, 2014.

[149] David T Limmer and David Chandler. Phase diagram of supercooled water confined to hydrophilic nanopores. *J. Chem. Phys.*, 137:044509, 2012.

[150] David T Limmer and David Chandler. Corresponding states for mesostructure and dynamics of supercooled water. *Faraday Discuss.*, 167:485–498, 2013.

[151] Thomas Speck and David Chandler. Constrained dynamics of localized excitations causes a non-equilibrium phase transition in an atomistic model of glass formers. *J. Chem. Phys.*, 136:184509, 2012.

[152] Erik Bitzek, Pekka Koskinen, Franz Gähler, Michael Moseler, and Peter Gumbsch. Structural relaxation made simple. *Phys. Rev. Lett.*, 97:170201, 2006.

[153] Paul J. Steinhardt, David R. Nelson, and Marco Ronchetti. Bond-orientational order in liquids and glasses. *Phys. Rev. B*, 28:784–805, 1983.

[154] Shankar Kumar, John M. Rosenberg, Djamal Bouzida, Robert H. Swendsen, and Peter A. Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.*, 13:1011–1021, 1992.

[155] Jochen S. Hub, Bert L. de Groot, and David van der Spoel. `g_wham` – a free weighted histogram analysis implementation including robust error and autocorrelation estimates. *J. Chem. Theor. Comput.*, 6:3713–3720, 2010.

[156] Eduardo Sanz, Chantal Valeriani, Emanuela Zaccarelli, W. C. K. Poon, P. N. Pusey, and M. E. Cates. Crystallization mechanism of hard sphere glasses. *Phys. Rev. Lett.*, 106:215701, 2011.

[157] Eduardo Sanz, Chantal Valeriani, Emanuela Zaccarelli, Wilson C. K. Poon, Michael E. Cates, and Peter N. Pusey. Avalanches mediate crystallization in a hard-sphere glass. *Proc. Natl. Acad. Sci. U.S.A*, 111:75–80, 2014.

[158] John Russo, Flavio Romano, and Hajime Tanaka. New metastable form of ice and its role in the homogeneous crystallization of water. *Nature Mater.*, 13:733–739, 2014.

[159] Kieron Burke. Perspective on density functional theory. *J. Chem. Phys.*, 136:150901, 2012.

[160] C. Fiolhais, F. Nogueira, and M. Marques. *A Primer in density functional theory*. Springer, Berlin [etc.], 2003.

[161] Robert G. Parr and Weitao Yang. *Density-Functional Theory of Atoms and Molecules*. Oxford University Press, USA, 1994.

[162] Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864, 1964.

[163] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, 1965.

[164] Jeffery B. Klauda and Stanley I. Sandler. Global distribution of methane hydrate in ocean sediment. *Energy and Fuels*, 19:459–470, 2005.

[165] Matthew R. Walsh, Carolyn A. Koh, E. Dendy Sloan, Amadeu K. Sum, and David T. Wu. Microsecond simulations of spontaneous methane hydrate nucleation and growth. *Science*, 326:1095–1098, 2009.

[166] Matthew R. Walsh, Gregg T. Beckham, Carolyn A. Koh, E. Dendy Sloan, David T. Wu, and Amadeu K. Sum. Methane hydrate nucleation rates from molecular dynamics simulations: effects of aqueous methane concentration, interfacial curvature, and system size. *J. Phys. Chem. C*, 115:21241–21248, 2011.

[167] Liam C. Jacobson, Waldemar Hujo, and Valeria Molinero. Amorphous precursors in the nucleation of clathrate hydrates. *J. Am. Chem. Soc.*, 132:11806–11811, 2010.

[168] Ravi Radhakrishnan and Bernhardt L. Trout. A new approach for studying nucleation phenomena using molecular simulations: Application to CO_2 hydrate clathrates. *J. Chem. Phys.*, 117:1786–1796, 2002.

[169] C. Moon, R. W. Hawtin, and P. Mark Rodger. Nucleation and control of clathrate hydrates: insights from simulation. *Faraday Discuss.*, 136:367–382, 2007.

[170] Stefan Grimme. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.*, 27:1787–1799, 2006.

[171] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.*, 132:154104, 2010.

[172] Alexandre Tkatchenko and Matthias Scheffler. Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data. *Phys. Rev. Lett.*, 102:073005, 2009.

[173] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist. Van der Waals density functional for general geometries. *Phys. Rev. Lett.*, 92:246401, 2004.

[174] Jiří Klimeš and Angelos Michaelides. Perspective: Advances and challenges in treating van der Waals dispersion forces in density functional theory. *J. Chem. Phys.*, 137:120901, 2012.

[175] Biswajit Santra, Jiři Klimeš, Dario Alfè, Alexandre Tkatchenko, Ben Slater, Angelos Michaelides, Roberto Car, and Matthias Scheffler. Hydrogen bonds and van der Waals forces in ice at ambient and high pressures. *Phys. Rev. Lett.*, 107:185701, 2011.

[176] Biswajit Santra, Jiří Klimeš, Alexandre Tkatchenko, Dario Alfè, Ben Slater, Angelos Michaelides, Roberto Car, and Matthias Scheffler. On the accuracy of van der Waals inclusive density-functional theory exchange-correlation functionals for ice at ambient and high pressures. *J. Chem. Phys.*, 139:154702, 2013.

[177] Brandon C. Knott, Valeria Molinero, Michael F. Doherty, and Baron Peters. Homogeneous nucleation of methane hydrates: Unrealistic under realistic conditions. *J. Am. Chem. Soc.*, 134:19544–19547, 2012.

[178] Annika Lenz and Lars Ojamäe. Structures of the I-, II- and H-methane clathrates and the ice- methane clathrate phase transition from quantum-chemical modeling with force-field thermal corrections. *J. Phys. Chem. A*, 115:6169–6176, 2011.

[179] Guillermo Román-Pérez, Mohammed Moaied, Jose M. Soler, and Felix Yndurain. Stability, adsorption, and diffusion of CH_4, CO_2, and H_2 in clathrate hydrates. *Phys. Rev. Lett.*, 105:145901, 2010.

[180] Qi Li, Brian Kolb, Guillermo Román-Pérez, José M. Soler, Felix Yndurain, Lingzhu Kong, D. C. Langreth, and T. Thonhauser. *Ab initio* energetics and kinetics study of H_2 and CH_4 in the sI clathrate hydrate. *Phys. Rev. B*, 84:153103, 2011.

[181] N. D. Drummond and R. J. Needs. Quantum Monte Carlo, density functional theory, and pair potential studies of solid neon. *Phys. Rev. B*, 73:024107, 2006.

[182] Luke Shulenburger and Thomas R. Mattsson. Quantum Monte Carlo applied to solids. *Phys. Rev. B*, 88:245117, 2013.

[183] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal. Quantum Monte Carlo simulations of solids. *Rev. Mod. Phys.*, 73:33–83, 2001.

[184] R. J. Needs, M. D. Towler, N. D. Drummond, and P. López Ríos. Continuum variational and diffusion quantum Monte Carlo calculations. *J. Phys.: Condens. Matter*, 22:023201, 2010.

[185] D. Alfè and M. J. Gillan. Efficient localized basis set for quantum Monte Carlo calculations on condensed matter. *Phys. Rev. B*, 70:161101, 2004.

[186] Biswajit Santra, Angelos Michaelides, Martin Fuchs, Alexandre Tkatchenko, Claudia Filippi, and Matthias Scheffler. On the accuracy of density-functional theory exchange-correlation functionals for H bonds in small water clusters. II. The water hexamer and van der Waals interactions. *J. Chem. Phys.*, 129:194111, 2008.

[187] M. J. Gillan, F. R. Manby, M. D. Towler, and D. Alfè. Assessing the accuracy of quantum Monte Carlo and density functional theory for energetics of small water clusters. *J. Chem. Phys.*, 136:244105, 2012.

[188] I. G. Gurtubay and R. J. Needs. Dissociation energy of the water dimer from quantum Monte Carlo calculations. *J. Chem. Phys.*, 127:124306, 2007.

[189] F. F. Wang, M. J. Deible, and K. D. Jordan. Benchmark study of the interaction energy of a $(H_2O)_{16}$ cluster. *J. Phys. Chem. B*, 117:7606–7611, 2013.

[190] Matúš Dubecký, Petr Jurečka, René Derian, Pavel Hobza, Michal Otyepka, and Lubos Mitas. Quantum Monte Carlo methods describe noncovalent

interactions with subchemical accuracy. *J. Chem. Theor. Comput.*, 9:4287–4292, 2013.

[191] Javier Carrasco, Biswajit Santra, Jiri Klimeš, and Angelos Michaelides. To wet or not to wet? Dispersion forces tip the balance for water ice on metals. *Phys. Rev. Lett.*, 106:026101, 2011.

[192] F. Mittendorfer, A. Garhofer, J. Redinger, J. Klimeš, J. Harl, and G. Kresse. Graphene on Ni(111): Strong interaction and weak adsorption. *Phys. Rev. B*, 84:201401, 2011.

[193] William L. Jorgensen, David S. Maxwell, and Julian Tirado-Rives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, 118:11225–11236, 1996.

[194] G. Kresse and J. Hafner. *Ab initio* molecular dynamics for liquid metals. *Phys. Rev. B*, 47:558–561, 1993.

[195] G Kresse and J Furthmüller. Efficiency of *ab-initio* total energy calculations for metals and semiconductors using a plane-wave basis set. *Comp. Mater. Sci.*, 6:15–50, 1996.

[196] G. Kresse and J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54:11169–11186, 1996.

[197] Guillermo Román-Pérez and José M. Soler. Efficient implementation of a van der Waals density functional: Application to double-wall carbon nanotubes. *Phys. Rev. Lett.*, 103:096102, 2009.

[198] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B*, 59:1758–1775, 1999.

[199] Hendrik J. Monkhorst and James D. Pack. Special points for Brillouin-zone integrations. *Phys. Rev. B*, 13:5188–5192, 1976.

[200] N. D. Drummond, M. D. Towler, and R. J. Needs. Jastrow correlation factor for atoms, molecules, and solids. *Phys. Rev. B*, 70:235119, 2004.

[201] Paolo Giannozzi, Stefano Baroni, Nicola Bonini, Matteo Calandra, Roberto Car, Carlo Cavazzoni, Davide Ceresoli, Guido L Chiarotti, Matteo Cococcioni, Ismaila Dabo, Andrea Dal Corso, Stefano de Gironcoli, Stefano Fabris, Guido Fratesi, Ralph Gebauer, Uwe Gerstmann, Christos Gougoussis, Anton Kokalj, Michele Lazzeri, Layla Martin-Samos, Nicola Marzari, Francesco Mauri, Riccardo Mazzarello, Stefano Paolini, Alfredo Pasquarello, Lorenzo Paulatto, Carlo Sbraccia, Sandro Scandolo, Gabriele Sclauzero, Ari P Seitsonen, Alexander Smogunov, Paolo Umari, and Renata M Wentzcovitch. `QUANTUM ESPRESSO`: a modular and open-source software project for quantum simulations of materials. *J. Phys.: Condens. Matter*, 21:395502, 2009.

[202] J. R. Trail and R. J. Needs. Smooth relativistic Hartree–Fock pseudopotentials for H to Ba and Lu to Hg. *J. Chem. Phys.*, 122:174109, 2005.

[203] J. R. Trail and R. J. Needs. Norm-conserving Hartree–Fock pseudopotentials and their asymptotic behavior. *J. Chem. Phys.*, 122:014112, 2005.

[204] Jie Ma, Dario Alfè, Angelos Michaelides, and Enge Wang. The water-benzene interaction: Insight from electronic structure theories. *J. Chem. Phys.*, 130:154303, 2009.

[205] Jie Ma, Angelos Michaelides, and Dario Alfè. Binding of hydrogen on benzene, coronene, and graphene from quantum Monte Carlo calculations. *J. Chem. Phys.*, 134:134701, 2011.

[206] S. Chiesa, D. M. Ceperley, R. M. Martin, and M. Holzmann. Finite-size error in many-body simulations with long-range interactions. *Phys. Rev. Lett.*, 97:076404, 2006.

[207] N. D. Drummond, R. J. Needs, A. Sorouri, and W. M. C. Foulkes. Finite-size errors in continuum quantum Monte Carlo calculations. *Phys. Rev. B*, 78:125106, 2008.

[208] A. J. Williamson, G. Rajagopal, R. J. Needs, L. M. Fraser, W.M.C. Foulkes, Y. Wang, and M. Y. Chou. Elimination of Coulomb finite-size effects in quantum many-body simulations. *Phys. Rev. B*, 55:R4851, 1997.

[209] L. M. Fraser, W. M. C. Foulkes, G. Rajagopal, R. J. Needs, S. D. Kenny, and A. J. Williamson. Finite-size effects and Coulomb interactions in quantum Monte Carlo calculations for homogeneous systems with periodic boundary conditions. *Phys. Rev. B*, 53:1814, 1996.

[210] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.*, 98:10089–10092, 1993.

[211] Ulrich Essmann, Lalith Perera, Max L. Berkowitz, Tom Darden, Hsing Lee, and Lee G. Pedersen. A smooth particle mesh Ewald method. *J. Chem. Phys.*, 103:8577–8593, 1995.

[212] R. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM J. Scientif. Comput.*, 16:1190–1208, 1995.

[213] Ciyou Zhu, Richard H. Byrd, Peihuang Lu, and Jorge Nocedal. Algorithm 778: `L-BFGS-B`: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Softw.*, 23:550–560, 1997.

[214] I-Ming Chou, Anurag Sharma, Robert C. Burruss, Jinfu Shu, Ho-kwang Mao, Russell J. Hemley, Alexander F. Goncharov, Laura A. Stern, and Stephen H. Kirby. Transformations in methane hydrates. *Proc. Natl. Acad. Sci. USA*, 97:13484–13487, 2000.

[215] H Hirai, T Tanaka, T Kawamura, Y Yamamoto, and T Yagi. Structural changes in gas hydrates and existence of a filled ice structure of methane hydrate above 40 GPa. *J. Phys. Chem. Solids*, 65:1555 – 1559, 2004.

[216] F. D. Murnaghan. The compressibility of media under extreme pressures. *Proc. Natl. Acad. Sci.*, 30:244, 1944.

[217] C. L. Fu and K. M. Ho. First-principles calculation of the equilibrium ground-state properties of transition metals: Applications to Nb and Mo. *Phys. Rev. B*, 28:5480–5486, 1983.

[218] D. W. Davidson, Y. P. Handa, C. I. Ratcliffe, J. S. Tse, and B. M. Powell. The ability of small molecules to form clathrate hydrates of structure-II. *Nature*, 311:142–143, 1984.

[219] C. Gutt, B. Asmussen, W. Press, M. R. Johnson, Y. P. Handa, and J. S. Tse. The structure of deuterated methane–hydrate. *J. Chem. Phys.*, 113:4713–4721, 2000.

[220] Jie Ma, Angelos Michaelides, Dario Alfè, Laurids Schimka, Georg Kresse, and Enge Wang. Adsorption and diffusion of water on graphene from first principles. *Phys. Rev. B*, 84:033402, 2011.

[221] Gabriella Graziano, Jiří Klimeš, Felix Fernandez-Alonso, and Angelos Michaelides. Improved description of soft layered materials with van der Waals density functional theory. *J. Phys.: Condens. Matter*, 24:424216, 2012.

[222] Y P Handa. Compositions, enthalpies of dissociation, and heat capacities in the range 85 to 270 K for clathrate hydrates of methane, ethane, and propane, and enthalpy of dissociation of isobutane hydrate, as determined by a heat-flow calorimeter. *J. Chem. Thermodynamics*, 18:915–921, 1986.

[223] Graydon K. Anderson. Enthalpy of dissociation and hydration number of methane hydrate from the Clapeyron equation. *J. Chem. Thermodynamics*, 36:1119–1127, 2004.

[224] Bertrand Guillot. A reappraisal of what we have learnt during three decades of computer simulations on water. *J. Mol. Liq.*, 101:219, 2002.

[225] J. L. F. Abascal, R. García Fernández, L. G. MacDowell, E. Sanz, and C. Vega. Ice: A fruitful source of information about liquid water. *J. Mol. Liq.*, 136:214, 2007.

[226] M. M. Conde and C. Vega. Determining the three-phase coexistence line in methane hydrates using computer simulations. *J. Chem. Phys.*, 133:064507, 2010.

[227] H. Jiang, K. D. Jordan, and C. E. Taylor. Molecular dynamics simulations of methane hydrate using polarizable force fields. *J. Phy. Chem. B*, 111:6486–6492, 2007.

[228] Michael J. Deible, Odbadrakh Tuguldur, and Kenneth D. Jordan. Theoretical study of the binding energy of a methane molecule in a $(H_2O)_{20}$ dodecahedral cage. *J. Phys. Chem. B*, 118:8257–8263, 2014.

[229] C Adamo and V Barone. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.*, 110:6158–6170, 1999.

[230] D. Alfè, A. P. Bartók, G. Csányi, and M. J. Gillan. Communication: Energy benchmarking with quantum Monte Carlo for water nano-droplets and bulk liquid water. *J. Chem. Phys.*, 138:221102, 2013.

[231] Albert P. Bartók, Mike C. Payne, Risi Kondor, and Gábor Csányi. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.*, 104:136403, 2010.

[232] Sheng Meng, E. G. Wang, and Shiwu Gao. Water adsorption on metal surfaces: A general picture from density functional theory studies. *Phys. Rev. B*, 69:195404, 2004.

[233] B Chen, J H Xing, and J I Siepmann. Development of polarizable water force fields for phase equilibrium calculations. *J. Phys. Chem. B*, 104:2391, 2000.

[234] Joost VandeVondele, Matthias Krack, Fawzi Mohamed, Michele Parrinello, Thomas Chassaing, and Jürg Hutter. Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.*, 167:103–128, 2005.

[235] G. Lippert, J. Hutter, and M. Parrinello. A hybrid Gaussian and plane wave density functional scheme. *Mol. Phys.*, 92:477, 1997.

[236] S. Goedecker, M. Teter, and J. Hutter. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B*, 54:1703–1710, 1996.

[237] C. Hartwigsen, S. Goedecker, and J. Hutter. Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Phys. Rev. B*, 58:3641–3662, 1998.

[238] Joost VandeVondele and Jürg Hutter. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J. Chem. Phys.*, 127:114105, 2007.

[239] J. Taylor and B Hale. Monte Carlo simulations of water-ice layers on a model silver iodide substrate: A comparison with bulk ice systems. *Phys. Rev. B*, 47:9732, 1993.

[240] Jessica C Johnston, Noah Kastelowitz, and Valeria Molinero. Liquid to quasicrystal transition in bilayer water. *J. Chem. Phys.*, 133:154516, 2010.

[241] Noah Kastelowitz, Jessica C. Johnston, and Valeria Molinero. The anomalously high melting temperature of bilayer ice. *J. Chem. Phys.*, 132:124511, 2010.

[242] Peter Hamm, Jan Helbing, and Jens Bredenbeck. Stretched versus compressed exponential kinetics in α-helix folding. *Chem. Phys.*, 323:54 – 65, 2006.

[243] Wolfgang Lechner and Christoph Dellago. Accurate determination of crystal structures based on averaged local bond order parameters. *J. Chem. Phys.*, 129:114707, 2008.

[244] John S. Tse, Michael L. Klein, and Ian R. McDonald. Computer simulation studies of the structure I clathrate hydrates of methane, tetrafluoromethane, cyclopropane, and ethylene oxide. *J. Chem. Phys.*, 81:6146–6153, 1984.

[245] Saman Alavi, J. A. Ripmeester, and D. D. Klug. Molecular dynamics study of the stability of methane structure H clathrate hydrates. *J. Chem. Phys.*, 126:124708, 2007.

[246] Wendy D. Cornell, Piotr Cieplak, Christopher I. Bayly, Ian R. Gould, Kenneth M. Merz, David M. Ferguson, David C. Spellmeyer, Thomas Fox, James W. Caldwell, and Peter A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, 117:5179–5197, 1995.

[247] A. D. MacKerell, D. Bashford, Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, 102:3586–3616, 1998.

[248] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman J.C Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.*, 23:327–341, 1977.

[249] E. S. Severin and D. J. Tildesley. A methane molecule adsorbed on a graphite surface. *Mol. Phys.*, 41:1401–1418, 1980.

[250] D. van der Spoel, E. Lindahl, B. Hess, A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. M. Sijbers, K. A. Feenstra, R. van Drunen, and H. J. C. Berendsen. *Gromacs User Manual version 4.5.4*, 2010.